A Sub-Additive DC Approach to the Complementarity Problem

Tangi Migot, Lina Abdallah, Mounir Haddou

Noname manuscript No.
(will be inserted by the editor)

A Sub-Additive DC Approach to the Complementarity Problem

L. Abdallah · M. Haddou · T. Migot

This paper has been accepted for publication in *Computational Optimization and Applications*, find the published version here: https://link.springer.com/article/10.1007/s10589-019-00078-w.

Abstract In this article, we study a merit function based on sub-additive functions for solving the non-linear complementarity problem (NCP). This leads to consider an optimization problem that is equivalent to the NCP. In the case of a concave NCP this optimization problem is a Difference of Convex (DC) program and we can therefore use DC Algorithm to locally solve it. We prove that in the case of a concave monotone NCP, it is sufficient to compute a stationary point of the optimization problem to obtain a solution of the complementarity problem. In the case of a general NCP, assuming that a DC decomposition of the complementarity problem is known, we propose a penalization technique to reformulate the optimization problem as a DC program and prove that local minima of this penalized problem are also local minima of the merit problem. Numerical results on linear complementarity problems, absolute value equations and non-linear complementarity problems show that our method is promising.

Keywords: complementarity problem - difference of convex - merit function - dc algorithm

AMS Subject Classification: 90C59 – 90C30 – 90C33 – 65K05 – 49M20

L. Abdallah
Lebanese University, Tripoli, Lebanon.
E-mail: lina_abdallah@hotmail.fr

M. Haddou
Univ Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France.
E-mail: mounir.haddou@insa-rennes.fr

T. Migot
University of Guelph, Department of Mathematics & Statistics, Canada.
E-mail: tangi.migot@gmail.com

1 Introduction

The non-linear complementarity problem (NCP) consists in finding $x \in \mathbb{R}^n$ satisfying

$$x \geq 0, F(x) \geq 0, x^T F(x) = 0, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be continuous. In the special case where F is an affine function, (1) is reduced to the linear complementarity problem (LCP).

The NCP has been very popular due to its numerous applications in economy, physics, chemistry ... see the monographs [5, 7] and references therein. This problem is a feasibility and not an optimization problem. However, it is well-known that it is closely related with optimization problems. Indeed, optimality conditions of many optimization problems may be cast as an NCP.

The family of methods to solve this problem can be divided in two: equation reformulation methods and merit function methods. In the former, the complementarity problem (1) is reformulated as an unconstrained (non-smooth) equation solved by a Newton-kind method or an unconstrained minimization problem depending on the assumptions on F, see [3, 4, 6, 13, 15, 22]. In the latter, the complementarity is ensured through a merit function reformulating (1) as an optimization problem, see [8, 9, 11, 17]. An extensive treatment of these methods and their extensions can be found in [7]. In the sequel, we focus on the merit function approach that is the most tractable option if we do not make any strong assumption on the problem.

A generic formulation of a complementarity problem reformulated as an optimization problem is given by

$$\min_{x \in \mathbb{R}^n} \Theta(x) \text{ s.t. } x \geq 0, F(x) \geq 0, \tag{2}$$

where $\Theta : \mathbb{R}^n \to \mathbb{R}$ is called a merit function if it satisfies the following property: x solves (1) if and only if x is the global minimizer of the previous problem. In this sense, we say that the optimization problem is equivalent to the complementarity problem. If the set of solutions of (1) is empty, then either the global optimal value of the objective function is positive or there is no global minima. It is to be noted that even so $F(x) = Mx + q$ the merit function, Θ, is not convex in general. Thus, it is fundamental to study the numerical methods used to solve the optimization problem (2). Due to the non-convexity, the realistic goal is to compute a local minimum or at least a stationary point of (2).

Our aim in this article is to use the very efficient local theory designed for difference of convex (DC) functions by considering a merit function that can be expressed as a sum of convex and concave functions. Local optimization approach for DC optimization and DC Algorithm (DCA) were introduced by Pham Dinh Tao in a preliminary form in the 80's [25] and extensively developed since then, see [23] for a review on this method. This approach has been widely used by many researchers, see a list of references on http://www.lita.univ-lorraine.fr/~lethi/index.php/dca.html. This work is motivated by the fact that

DCA has been successfully applied to many (smooth or non-smooth) large-scale non-convex programs in various domains of applied sciences. In a recent article [17], the authors present four formulations of the LCP as a DC optimization problem and show promising numerical results.

Motivated by the recent development of DC approaches to solve the complementarity problems, we study a merit function based on a family of smooth sub-additive functions that has been used in the literature in the context of sparse optimization [20] and absolute value equations [1], whose latter can be reformulated as an LCP.

One of our main results considering this merit function is that stationary points of (2) are solutions of a concave monotone NCP. Therefore, in this favorable case, there is no need to study global minima of the problem and this proves that our approach is not more difficult than equation based approach in this advantageous cases.

Besides, we prove that the optimization problem (2) with our merit function is a DC program if F is concave. This property permits us to use the local theory to solve DC program and in particular DCA. In the case of the general NCP, assuming a DC composition of F is known, the optimization problem with our merit function can be reformulated as a DC program through penalizations. We prove equivalence between local minima of this penalized DC program and the local minima of (2).

Numerical results on linear complementarity problems, absolute value equations and non-linear complementarity problems show the interest of the proposed method.

In Section 2, we introduce the merit function that is proved to be theoretically sound for the monotone NCP in Section 3. In Section 4, we show that this new formulation leads to a DC program that can be solved using DCA as explained in the Section 5. Finally, in Section 6, we give several numerical experiments to show the validity of our method compared to existing methods in the literature on linear and non-linear problems.

2 A Merit Function for Complementarity Problems

In this section, we present a reformulation of the complementarity problem through a merit function. This function has been first used in [1] as a heuristic to solve absolute value equations reformulated as mixed linear complementarity problems.

Let us consider the merit function $\Theta : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$ defined by

$$\Theta(x, F(x)) := \sum_{i=1}^{n} \theta(x_i) + \theta(F_i(x)) - \theta(x_i + F_i(x)), \tag{3}$$

where functions $\theta : \mathbb{R}^n_+ \to [0, 1[$ are non-decreasing strictly concave continuously differentiable functions such that $\theta(0) = 0$. Examples of these functions for $r > 0$ are:

$$\begin{aligned}
&(\theta^1): \quad \theta(x) = \frac{x}{x+r},\\
&(\theta^2): \quad \theta(x) = \left(1 - e^{-(\frac{x}{r})}\right)^k, \text{ for } k \geq 1,\\
&(\theta^{log}): \theta(x) = \frac{\log(1+x)}{\log(1+x+r)}.
\end{aligned}$$

For r sufficiently small, the functions θ approximate a step function. This observation already leads to interesting results in sparse optimization in [20] and in the context of complementarity problems in [1,13].

We use the function Θ to reformulate the NCP as follows

$$\Theta(x, F(x)) = 0, \; x \geq 0, \; F(x) \geq 0. \tag{4}$$

The following lemma shows that the functions θ are sub-additive.

Lemma 2.1 *For all $x \geq 0$ such that $F(x) \geq 0$, we have*

$$\forall i, \; \theta(x_i) + \theta(F_i(x)) \geq \theta(x_i + F_i(x)),$$

with equality if and only if $x_i = 0$ or $F_i(x) = 0$.

Proof Since θ is concave, we obtain

$$\forall z \neq y \in \mathbb{R}, \forall t \in (0,1), \; t\theta(z) + (1-t)\theta(y) \leq \theta(tz + (1-t)y),$$

with equality if $t = 0$ or 1 and if $z = y$. For $y = 0$ and $\theta(0) = 0$ yields to

$$\theta(tz) = \theta(tz + (1-t)y) \geq t\theta(z), \; \forall t \in (0,1),$$

with equality if $t = 0, 1$ or $z = 0$. Take $i \in \{1, ..., n\}$, $z = x_i$, $y = F_i(x)$ and suppose that $x_i + F_i(x) \neq 0$ (the case $x_i = F_i(x) = 0$ stay true)

$$\begin{aligned}
\theta(x_i) + \theta(F_i(x)) &= \theta\left((x_i + F_i(x))\frac{x_i}{x_i + F_i(x)}\right) + \theta\left((x_i + F_i(x))\frac{F_i(x)}{x_i + F_i(x)}\right),\\
&\geq \frac{x_i}{x_i + F_i(x)}\theta((x_i + F_i(x))) + \frac{F_i(x)}{x_i + F_i(x)}\theta((x_i + F_i(x))),\\
&= \theta(x_i + F_i(x)),
\end{aligned}$$

with equality if and only if $x_i = 0$ or $F_i(x) = 0$. $\square$

It is now straightforward to conclude that finding a global minimum of (4) is equivalent to solving (1).

Theorem 2.1 *Let $x \in \mathbb{R}^n_+$ such that $F(x) \in \mathbb{R}^n_+$. It holds that*

$$\Theta(x, F(x)) = 0 \Longleftrightarrow x^T F(x) = 0.$$

Proof By Lemma 2.1, Θ is a sum of non-negative terms, therefore Θ is equal to 0 if and only if all terms are equal to zero. In addition, for a component i there is equality if and only if $x_i = 0$ or $F_i(x) = 0$. $\square$

Using (3) to define a sub-additive merit function, we consider the following optimization problem:

$$\inf_{x \in \mathbb{R}^n} \Theta(x, F(x)) \text{ s.t. } x \geq 0, F(x) \geq 0. \tag{P}$$

We denote $\mathcal{F}$ the feasible set of (P) defined as

$$\mathcal{F} := \{x \in \mathbb{R}^n \mid x \geq 0, F(x) \geq 0\}.$$

It is to be noted that the feasible set of (P) could be non-convex, unless the function F is assumed concave. The following result that is given without proof sums up the possible situations that may arise by solving (P) and their consequences to our initial aim that is solving (1).

Theorem 2.2 *Assume that $F : \mathbb{R}^n_+ \to \mathbb{R}^n$ is a continuous function and that $\mathcal{F}$ is non-empty. Moreover, assume that there is no sequence $\{x^k\} \in \mathcal{F}$ such that $\lim_{\|x^k\| \to \infty} F(x^k)^T x^k = 0$. Then, exactly one of the following cases arises:*

- *The global optimal value of* (P) *is zero and the corresponding point is a solution of* (1)*;*
- *The global optimal value of* (P) *is positive and* (1) *has no solution.*

An assumption similar to the one given in this theorem was used in Corollary 2.6.2 in [7] to guarantee existence of a solution of the (NCP).

3 Sub-Additive Reformulation for the Monotone NCP

In this section, we prove that for a popular class of NCPs finding a stationary point of (P) is actually sufficient to solve the initial problem. First, let us introduce the monotone NCP.

3.1 Monotone Complementarity Problems

We focus on NCPs that are said monotone, i.e. F is P function (see definition below). We give classical definitions and characterization of this family of complementarity problems.

Definition 3.1 *[7] The function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a P_0 function iff*

$$\max_{i: x_i \neq y_i} (x - y)_i (F_i(x) - F_i(y)) \geq 0 \quad \forall x \neq y \in \mathbb{R}^n.$$

If this inequality is strict, F is said to be a P function.

Assuming differentiability of the function F this definition reduces to a condition on the jacobian matrix. We extend the definition of a P_0 (resp. P) function to a P_0 (resp. P) matrix.

Definition 3.2 *[7] A matrix $M \in \mathbb{R}^{n \times n}$ is said to be a:*

– P_0 *matrix if all of its principal minors are nonnegative;*
– P *matrix if all of its principal minors are positive;*
– *positive semidefinite if* M *is symmetric and* $x^T Mx \geq 0$ *for all* $x \in \mathbb{R}^n$*, and positive definite if* $x^T Mx > 0$ *for all* $x \in \mathbb{R}^n (x \neq 0)$.

A P matrix is a generalization of a positive definite matrix, while a P_0 matrix is a generalization of a positive semi-definite matrix. In the case of a smooth function F, then Proposition 3.5.9 of [7] gives the following characterization.

Proposition 3.1 *For a continuously differentiable function* F*, if its transpose jacobian* $\nabla F(x)^T$ *is a* P *matrix then* F *is a* P *function, whereas the transpose jacobian* $\nabla F(x)^T$ *is a* P_0*-matrix if and only if* F *is a* P_0 *function.*

The monotone NCPs are particularly useful since they provide a characterization of the set of solutions of the NCP. The following proposition is a sum up of results given in [7].

Proposition 3.2 *If* F *is a* P *function, then* (1) *has at most one solution.*

Among other interesting properties, we show in Lemma 3.1 that some regularity may be obtained for the set $\mathcal{F}$ for this family of NCPs. Finally, the following result from Theorem 3.3.4,[5] gives another characterization of P-matrix.

Proposition 3.3 *Let* $M \in \mathbb{R}^{n\times n}$*.* M *is a P-matrix if and only if*

$$z_i(Mz)_i \leq 0 \text{ for all } i \Longrightarrow z = 0.$$

3.2 A Sufficient Optimality Condition for the Montone NCP

We assume here that the function F is a continuously differentiable function. A classical approach in non-linear programming is to study stationary points of (P) using the classical Karush-Kuhn-Tucker conditions. Let $x^* \in \mathcal{F}$ be a stationary point of (P) if there exists $\lambda, \mu \in \mathbb{R}^n \times \mathbb{R}^n$ satisfying

$$\begin{array}{l}\nabla \Theta(x^*, F(x^*)) - \nabla F(x^*)^T \lambda - \mu = 0,\\ \lambda^T F(x^*) = 0, F(x^*) \geq 0, \lambda \geq 0,\\ \mu^T x^* = 0, x^* \geq 0, \mu \geq 0.\end{array}$$

It holds from non-linear programming theory that a local minimum of (P) that satisfies a constraint qualification is a stationary point of (P). Let us now give two examples of classical constraint qualifications. Beforehand, we use the following notations:

$$\begin{aligned}\mathcal{I}_F(x) &:= \{i = 1, \dots, n \mid F_i(x) = 0\}, \text{ and,}\\ \mathcal{I}(x) &:= \{i = 1, \dots, n \mid x_i = 0\}.\end{aligned}$$

Definition 3.3 *We say that* $x^* \in \mathcal{F}$ *satisfies*

– *the Linear Independence CQ (LICQ) if the gradients*

$$\{\nabla F_i(x^*) \mid \forall i \in \mathcal{I}_F(x^*)\} \cup \{\nabla x_i^* \mid \forall i \in \mathcal{I}(x^*)\}$$

are linearly independent.

– *the Positive Linear Independence CQ (PLICQ) if the only solution to the equation*

$$0 = \sum_{i\in\mathcal{I}} \mu_i \nabla x_i^* + \sum_{i\in\mathcal{I}_F} \lambda_i \nabla F_i(x^*)$$

with $(\mu, \lambda) \geq 0$ is the trivial solution. We say in this case that the gradients

$$\{\nabla F_i(x^*) \mid \forall i \in \mathcal{I}_F(x^*)\} \cup \{\nabla x_i^* \mid \forall i \in \mathcal{I}(x^*)\}$$

are positively linearly independent.

It is known from [21] that PLICQ is equivalent to Mangasarian-Fromowitz CQ (MFCQ).

In the special case of a monotone NCP, the following lemma shows that some constraint qualification holds for the set $\mathcal{F}$.

Lemma 3.1 *Let F be a continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}^n$. Assume that for all $x \in \mathbb{R}^n$, $\nabla F(x)^T$ is a P matrix. Then, PLICQ holds at any point $x^* \in \mathcal{F}$. Furthermore, if x^* satisfies $x^* + F(x^*) > 0$, then LICQ holds at any point $x^* \in \mathcal{F}$.*

The assumption that $x^* + F(x^*) > 0$ is also well-known in the literature of complementarity problems under the name of strict complementarity.

In this proof, I_n denotes the identity matrix of size $n \times n$, and denotes a submatrix of the jacobian matrix by $\nabla F_{I\times J} := (\nabla F_{ij})_{i\in I, j\in J}$. Since the sets of indices are only evaluated at the point x^* we omit the dependence, i.e. $\mathcal{I} = \mathcal{I}(x^*)$ and $\mathcal{I}_F = \mathcal{I}_F(x^*)$.

Proof First, let us prove that given $x^* \in \mathcal{F}$ if x^* satisfies $x^* + F(x^*) > 0$, then LICQ holds at x^*. Let $(\lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^n$ be such that

$$0 = \sum_{i\in\mathcal{I}} \mu_i \nabla x_i^* + \sum_{i\in\mathcal{I}_F} \lambda_i \nabla F_i(x^*). \tag{5}$$

Since x^* satisfies $x^* + F(x^*) > 0$, then (5) is a combination of at most n gradients. We arrange the components of x and $F(x)$ such that $x_i F_i(x) = 0$ for $i = 1, \dots, m$ and $x_i > 0, F_i(x) > 0$ for $i = m+1, \dots, n$, so we can rewrite the m first equations of (5) as

$$\begin{pmatrix} I_{|\mathcal{I}|} & \nabla F_{\mathcal{I}\times\mathcal{I}_F}(x^*)^T \\ 0 & \nabla F_{\mathcal{I}_F\times\mathcal{I}_F}(x^*)^T \end{pmatrix} \begin{pmatrix} \mu_{\mathcal{I}} \\ \lambda_{\mathcal{I}_F} \end{pmatrix} = 0.$$

The jacobian matrix of F is a P matrix and in particular all its principal minor are non-singular matrices. Computing the Schur complement on the left-hand side it holds that the matrix is non-singular, therefore $\mu_{\mathcal{I}} = \lambda_{\mathcal{I}_F} = 0$.

So, LICQ holds at x^*.

Consider the case where there exists an index i such that $x_i^* + F_i(x^*) = 0$. Obviously (5) may be a combination of more than $n+1$ gradients and so, LICQ is unlikely to hold. However, we prove that PLICQ holds at x^*. Assume that $(\lambda, \mu) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$. We can rewrite the equations in $\mathcal{I}_F$ as

$$\nabla F_{\mathcal{I}_F \times \mathcal{I}_F}(x^*)^T \lambda_{\mathcal{I}_F} = - \begin{pmatrix} \mu_{\mathcal{I} \cap \mathcal{I}_F} \\ 0 \end{pmatrix}.$$

Since $\mu \geq 0$ and $\lambda \geq 0$ by assumption, Proposition 3.3 gives that $\lambda_{\mathcal{I}_F} = \mu_{\mathcal{I} \cap \mathcal{I}_F} = 0$. Then, it follows that $\mu_{\mathcal{I}} = 0$. So, the gradients are positively linearly independent and the result follows. □

The previous lemma guarantees that any local minimum of (P) is a stationary point. The following theorem proves that computing a stationary point of (P) may be sufficient to solve (1).

Theorem 3.1 *Let F be a continuously differentiable and assume that for all $x \in \mathbb{R}^n$, $\nabla F(x)^T$ is a P matrix. Suppose that a solution of* (1) *exists. Then, any KKT point of* (P) *that satisfies the strict complementarity condition, is a solution of* (1).

Proof Through the proof, we denote $\theta^{'}(x)$ for $x \in \mathbb{R}^n$, the vector of componentwise derivatives of θ, i.e. $\theta^{'}(x) := (\theta^{'}(x_i))_{1 \leq i \leq n}$.

Let $x \in \mathcal{F}$ be a stationary point of (P). Thus, there exists $\lambda, \mu \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ such that

$$\begin{aligned}
0 = &\theta^{'}(x) + \nabla F(x)^T \theta^{'}(F(x)) - \theta^{'}(x + F(x)) - \nabla F(x)^T \theta^{'}(x + F(x)) \\
&- \nabla F(x)^T \lambda - \mu, \\
&\lambda^T F(x) = 0, F(x) \geq 0, \lambda \geq 0, \\
&x^T \mu = 0, x \geq 0, \mu \geq 0.
\end{aligned}$$

We show the result by contradiction. Assume that x is not a solution of the NCP, then there exists a set of indices $J \neq \emptyset$ such that

$$J := \{i = 1, \dots, n \mid x_i > 0 \, \text{and} \, F_i(x) > 0\}.$$

Denote the sets $I_1 = \{j = 1, \dots, n \mid x_j = 0, F_j(x) > 0\}, I_2 = \{k = 1, \dots, n \mid x_k > 0, F_k(x) = 0\}$. By the strict complementarity assumption, it always holds that $I_1 \cup J \cup I_2 = \{1, \dots, n\}$.

Rewriting the previous equation yields

$$\begin{aligned}
\begin{pmatrix} 0_{I_1} \\ 0_J \\ 0_{I_2} \end{pmatrix} &= \begin{pmatrix} \theta^{'}(0) \\ \theta^{'}(x_J) \\ \theta^{'}(x_{I_2}) \end{pmatrix} - \begin{pmatrix} \theta^{'}(F_{I_1}(x)) \\ \theta^{'}(x_J + F_J(x)) \\ \theta^{'}(x_{I_2}) \end{pmatrix} + \nabla F(x)^T \begin{pmatrix} \theta^{'}(F_{I_1}(x)) \\ \theta^{'}(F_J(x)) \\ \theta^{'}(0) \end{pmatrix} \\
&\quad - \nabla F(x)^T \begin{pmatrix} \theta^{'}(F_{I_1}(x)) \\ \theta^{'}(F_J(x) + x_J) \\ \theta^{'}(x_{I_2}) \end{pmatrix} - \nabla F(x)^T \begin{pmatrix} 0_{I_1} \\ 0_J \\ \lambda_{I_2} \end{pmatrix} - \begin{pmatrix} \mu_{I_1} \\ 0_J \\ 0_{I_2} \end{pmatrix}, \\
&= \begin{pmatrix} \theta^{'}(0) - \theta^{'}(F_{I_1}(x)) \\ \theta^{'}(x_J) - \theta^{'}(x_J + F_J(x)) \\ 0 \end{pmatrix} - \begin{pmatrix} \mu_{I_1} \\ 0_J \\ 0_{I_2} \end{pmatrix} \\
&\quad + \nabla F(x)^T \begin{pmatrix} 0 \\ \theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x)) \\ \theta^{'}(0) - \theta^{'}(x_{I_2}) - \lambda_{I_2} \end{pmatrix}.
\end{aligned}$$

The equation on I_2 gives

$$\begin{aligned}
0 &= \nabla F_{I_2 \times J}(x)^T (\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x))) \\
&\quad + \nabla F_{I_2 \times I_2}(x)^T (\theta^{'}(0) - \theta^{'}(x_{I_2}) - \lambda_{I_2}), \\
\lambda_{I_2} &= (\nabla F_{I_2 \times I_2}(x)^T)^{-1} \nabla F_{I_2 \times J}(x)^T (\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x))) + \theta^{'}(0) - \theta^{'}(x_{I_2}).
\end{aligned}$$

Existence of the inverse of the matrix $\nabla F_{I_2 \times I_2}(x)^T$ is given by the jacobian matrix of F being a P matrix.

Replacing λ_{I_2} in the equation on J yields to

$$\begin{aligned}
0 &= \theta^{'}(x_J) - \theta^{'}(x_J + F_J(x)) + \nabla F_{J \times J}(x)^T (\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x))) \\
&\quad + \nabla F_{J \times I_2}(x)^T (\theta^{'}(0) - \theta^{'}(x_{I_2}) - \lambda_{I_2}), \\
0 &= \theta^{'}(x_J) - \theta^{'}(x_J + F_J(x)) + C(\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x))).
\end{aligned}$$

The matrix $C := \nabla F_{J \times J}(x)^T - \nabla F_{J \times I_2}(x)^T (\nabla F_{I_2 \times I_2}(x)^T)^{-1} \nabla F_{I_2 \times J}(x)^T$ corresponds to the Schur complement of the matrix $\nabla F_{I_2+J \times I_2+J}(x)^T$, whose determinant is positive by the P function assumption. Indeed, the determinant of the Schur complement satisfies

$$\det(\nabla F_{I_2+J \times I_2+J}(x)^T) = \det(\nabla F_{I_2 \times I_2}(x)^T) \det(C),$$

where, by the P function assumption, $\det(\nabla F^T_{I_2+J \times I_2+J})$ and $\det(\nabla F_{I_2 \times I_2}(x)^T)$ are both positive.

Furthermore, using the same formula, it can be verified that the principal minors of C have also a positive determinant. So, C is P-matrix.

Now, by Proposition 3.3, it follows that

$$C\left(\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x))\right)$$

has at least one positive component, since $\theta^{'}(F_J(x)) - \theta^{'}(x_J + F_J(x)) > 0$ by strict concavity of θ. However, this leads to a contradiction with the last equation, since strict concavity of θ also gives that $\theta^{'}(x_J) - \theta^{'}(x_J + F_J(x)) > 0$. So, this contradicts the existence of the set J and the result follows. □

We present here a counter-example to show that we cannot reduce the P function assumption to P_0 function.

Example 3.1 *Let* $F(x) = Mx + q$ *such that*

$$q = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \text{ and } M = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

M is a P_0 matrix. The point $x^* = (1,1)^T$ is clearly not a solution of the NCP, since $F(x^*) = Mx^* + q = (0,1)^T$. However it is a stationary point of (P) *with* $\lambda_2 = \mu_1 = \mu_2 = 0$ *and*

$$\lambda_1 = \theta'(1) + \theta'(0) - 2\theta'(2).$$

Following the notations in the previous proof, we have $I_1 = \emptyset, J = 2$ and $I_2 = 1$. The equation on I_2 gives

$$\nabla F_{I_2 \times J}(x^*)^T (\theta^{'}(F_J(x^*)) - \theta^{'}(x_J^* + F_J(x^*))) = -\nabla F_{I_2 \times I_2}(x^*)^T (\theta^{'}(0) - \theta^{'}(x_{I_2}^*) - \lambda_{I_2}),$$

since $\nabla F_{I_2 \times J}(x^*)^T = \nabla F_{I_2 \times I_2}(x^*)^T = 0$. *The equation on J gives*

$$\begin{aligned} 0 &= \theta^{'}(x_J^*) - \theta^{'}(x_J^* + F_J(x^*)) + \nabla F_{J \times J}(x^*)^T (\theta^{'}(F_J(x^*)) - \theta^{'}(x_J^* + F_J(x^*))) \\ &\quad + \nabla F_{J \times I_2}(x^*)^T (\theta^{'}(0) - \theta^{'}(x_{I_2}^*) - \lambda_{I_2}), \\ &= \theta'(1) - \theta'(2) + \theta'(1) - \theta'(2) + \theta'(0) - \theta'(1) - \lambda_1, \end{aligned}$$

which holds true by definition of λ_1.

We now present another affine example, where M is not a P matrix but a symmetric negative definite one.

Example 3.2 *Let* $F(x) = Mx + q$ *such that*

$$q = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \text{ and } M = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In this case, we have

$$\Theta(x, F(x)) = \theta(x_1) + \theta(x_2) + \theta(-x_1 + 2) + \theta(-x_2 + 2) - \theta(2) - \theta(2).$$

Computing the gradient of the objective function equal to zero yields

$$\begin{aligned} \theta'(x_1) - \theta'(-x_1 + 2) = 0, \\ \theta'(x_2) - \theta'(-x_2 + 2) = 0. \end{aligned}$$

The point $x^* = (1,1)^T$ satisfies $\nabla\Theta(x^*, F(x^*)) = 0$ and therefore is a stationary point of (P). *However, this point is not a solution of* (1).

An open question to this fundamental result is the possibility to remove the strict complementarity assumption. We left this question to future research.

4 A DC Formulation of the Sub-Additive Merit Function for the NCP

A function f is called a DC (difference of convex) function if it can be expressed as a difference of convex functions $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ such that

$$f = g - h.$$

The definition remains true if g and h are concave functions. A so-called DC program is that of minimizing a DC function over a convex set. A DC function has infinitely many decomposition g, h but the numerical resolution may be sensitive to this choice.

We split this section in two parts. First, we consider the formulation (P) in the case where F is a concave function. Second, we focus on the general case where we only assume that a DC decomposition of the function F is known.

4.1 The Concave NCP

In this section, we show that if F is a concave function the sub-additive function (3) is a DC function and the optimization problem (P) is a DC program. First, let us prove a simple lemma that shows that the composition of concave functions does not alter the concavity.

Lemma 4.1 *Let f be a concave function on $\mathbb{R}^n$, and let ϕ be a non-decreasing concave function on $\mathbb{R}^n$. Then, $h(x) = \phi(f(x))$ is a concave function.*

Proof For $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$, we have

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y).$$

Applying ϕ to both sides of the inequality, we get

$$h((1-\lambda)x + \lambda y) \geq \phi\left((1-\lambda)f(x) + \lambda f(y)\right).$$

By the concavity of the function ϕ it follows

$$h((1-\lambda)x + \lambda y) \geq \phi\left((1-\lambda)f(x) + \lambda f(y)\right) \geq (1-\lambda)h(x) + \lambda h(y).$$

Thus, h is concave. $\square$

We now prove that the objective function of (P) is a DC function.

Theorem 4.1 *Let F be a concave function on $\mathbb{R}^n_+$. Then,* (P) *is a DC program.*

Proof It is clear by concavity hypothesis on F that the set of constraints of (P) is a convex set. Now, let us verify that the function $\Theta(x, F(x))$ is a DC function with

$$g(x) = -\sum_{i=1}^{n} \theta(x_i + F_i(x)) \text{ and } h(x) = -\sum_{i=1}^{n} \theta(x_i) + \theta(F_i(x)).$$

Since F is a concave function and θ is a non-decreasing concave function we can apply Lemma 4.1 to conclude that g and h are both convex. □

It is to be noted that the hypothesis to take F concave is not so restrictive, since it includes the large family of LCPs.

4.2 The General NCP

In the special case where a DC decomposition of the function F is known we can reformulate (P) as a DC program. Given two convex functions $g, h : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$F(x) = g(x) - h(x).$$

We can reformulate (P) as

$$\begin{aligned} &\inf_{x,y \in \mathbb{R}^n} \Theta(x, y) \\ &\text{s.t. } x \geq 0, y \geq 0, \\ &\quad\quad y = F(x). \end{aligned}$$

Using the fact that a DC decomposition of F is known the previous problem becomes

$$\begin{aligned} &\inf_{x,y \in \mathbb{R}^n} \Theta(x, y) \\ &\text{s.t. } x \geq 0, y \geq 0, \\ &\quad\quad y = g(x) - h(x). \end{aligned}$$

Even so we use a DC decomposition of F and the objective function is a DC function, the set of constraints in the previous formulation remains non-convex. To tackle this problem, we try to remove the non-convexity in the constraints using slack variables and then penalizing the remaining non-convex constraints. Therefore, we consider the following penalized problem

$$\begin{aligned} &\inf_{x,y,t \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n} \Psi(x, y, t) := \Theta(x, y) + \tau \mathbf{e}^T (t - y - h(x)) + \tau \mathbf{e}^T (t - g(x)) \\ &\quad\quad \text{s.t. } (x, y, t) \in C, \\ &\quad\quad\quad\quad t \geq g(x),\ t - h(x) \geq y, \end{aligned} \tag{Pen_τ}$$

where $C := \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$. The link between the initial problem (P) and the penalized problem (Pen_τ) will be studied in Theorem 4.3 and 4.4. First, it is straightforward to see that the penalized problem is a DC program.

Theorem 4.2 *Given convex functions $g, h : \mathbb{R}^n \to \mathbb{R}^n$. Then, (Pen_τ) is a DC program.*

Proof We already shown in Theorem 4.1 that the function $\Theta(x, F(x))$ is a DC function if F is a concave function. However, it is to be noted here that we consider $\Theta(x, y)$, which is therefore a DC function.

Besides by convexity of the functions g and h it follows that the penalty terms in the objective function are concave. Thus, the objective function of (Pen_τ) is a DC function.

A similar argument gives that the constraints of (Pen_τ) are convex constraints. This completes the proof. □

We now investigate the link between the penalized DC program (Pen_τ) and the original problem (P).

Lemma 4.2 *Assume that $\mathcal{F}$ is non-empty. Let (x_τ, y_τ, t_τ) be a solution of (Pen_τ) and $\bar{x}$ be a feasible point of* (P)*, i.e. $\bar{x} \in \mathcal{F}$. Then, for all non-negative τ it holds that*

$$0 \le \Psi(x_\tau, y_\tau, t_\tau) \le \Theta(\bar{x}, F(\bar{x})).$$

Proof Since $\bar{x} \in \mathcal{F}$, it holds that $(\bar{x}, F(\bar{x}), g(\bar{x}))$ is a feasible point for (Pen_τ). Thus, by definition of the triple (x_τ, y_τ, t_τ) we have

$$\Psi(x_\tau, y_\tau, t_\tau) \le \Theta(\bar{x}, F(\bar{x})) + \tau \mathbf{e}^T(g(\bar{x}) - F(\bar{x}) - h(\bar{x})) + \tau \mathbf{e}^T(g(\bar{x}) - g(\bar{x})),$$

and since $F(\bar{x}) = g(\bar{x}) - h(\bar{x})$ we get

$$\Psi(x_\tau, y_\tau, t_\tau) \le \Theta(\bar{x}, F(\bar{x})).$$

The non-negativity of the left hand-side is a consequence of Lemma 2.1 and that (x_τ, y_τ, t_τ) is feasible for (Pen_τ). This completes the proof. □

A direct consequence of this lemma is equivalence of global minima of the penalized DC program (Pen_τ) and the original problem (P).

Theorem 4.3 *Assume that* (1) *has a bounded solution. Then, x^* is a global solution of* (P) *if and only if there exists y^*, t^* such that (x^*, y^*, t^*) is a global solution of (Pen_τ).*

Proof Let x^* be a solution of (P). Such solution exists, since we assume that (1) has at least one bounded solution. Then, the triple $(x^*, F(x^*), g(x^*))$ is feasible for (Pen_τ) and satisfies

$$\Theta(x^*, F(x^*)) + \tau \mathbf{e}^T(g(x^*) - F(x^*) - h(x^*)) + \tau \mathbf{e}^T(g(x^*) - g(x^*)) = 0,$$

by Theorem 2.1. This proves one side of the equivalence.

Now, let (x_τ, y_τ, t_τ) be a global solution (Pen_τ). Then, by Lemma 4.2 we have

$$0 \le \Theta(x_\tau, y_\tau) + \tau \mathbf{e}^T(t_\tau - y_\tau - h(x_\tau)) + \tau \mathbf{e}^T(t_\tau - g(x_\tau)) \le \Theta(\bar{x}, F(\bar{x})),$$

for all $\bar{x} \in \mathcal{F}$. In particular, consider $\bar{x}$ to be a solution of (1), then $\bar{x} \in \mathcal{F}$ and by Theorem 2.1 we get

$$0 = \Theta(x_\tau, y_\tau) + \tau \mathbf{e}^T(t_\tau - y_\tau - h(x_\tau)) + \tau \mathbf{e}^T(t_\tau - g(x_\tau)).$$

Noticing that the right hand-side is a sum of positive terms yields
$0 = \Theta(x_\tau, y_\tau)$,$y_\tau = F(x_\tau)$ and $t_\tau = g(x_\tau)$. By Theorem 2.1 it follows that x_τ is a solution of (1) and also a solution of (P). This completes the proof. □

We now focus on local minima of the penalized formulation. We notice, as proved in the following lemma, that the function Ψ is non-increasing with respect to the second variable.

Lemma 4.3 *Let $\tau > \theta'(0)$ and $x \geq 0$. Then, the function $\mathbb{R}_+ \ni y \mapsto \theta(y) - \theta(x+y) - \tau y$ is a non-increasing function.*

Proof Computing the derivative with respect to y of the function $\mathbb{R}_+ \ni y \mapsto \theta(y) - \theta(x+y) - \tau y$ gives

$$\theta'(y) - \theta'(x+y) - \tau.$$

However, this expression is negative since the functions θ are non-decreasing and by the concavity $\theta'(y) \leq \theta'(0) < \tau$. □

The following result considers the case of local optimality and shows that computing a local minimum of the penalized formulation is sufficient to find a local minimum of the initial formulation.

Theorem 4.4 *Let $\tau > \theta'(0)$. For any local minimum (x_τ, y_τ, t_τ) of (Pen_τ), then x_τ is a local minimum of* (P).

Proof Let (x_τ, y_τ, t_τ) be a local minimum of (Pen_τ), i.e. there exists $\delta > 0$ such that for all (x, y, t) in the ball of radius δ centered in (x_τ, y_τ, t_τ), denoted $\mathcal{B}_\delta(x_\tau, y_\tau, t_\tau)$, it holds that

$$\Psi(x_\tau, y_\tau, t_\tau) \leq \Psi(x, y, t).$$

By construction, it is clear that any feasible point of the problem (P) is also feasible for the problem (Pen_τ). Thus, it follows that if (x_τ, y_τ, t_τ) satisfies

$$t_\tau = g(x_\tau) \text{ and } y_\tau = g(x_\tau) - h(x_\tau), \tag{6}$$

then x_τ is a local minimum of (P).

Assume by contradiction that expression in (6) is wrong. In other words, there exists sets $T_t \subset \{1, \dots, n\}$ and $T_y \subset \{1, \dots, n\}$ such that for all $i \in T_t$ and all $j \in T_y$ we have

$$t_{\tau,i} > g_i(x_\tau) \text{ and } y_{\tau,j} < t_{\tau,j} - h_j(x_\tau). \tag{7}$$

Let us consider a point $(x_\tau, \bar{y}, \bar{t})$ such that

$$\bar{t} := \begin{cases} t_{\tau,i} - \bar{\delta}/2, & \text{if } i \in T_t, \\ t_{\tau,i}, & \text{otherwise,} \end{cases}$$

and

$$\bar{y} := \begin{cases} y_{\tau,j} + \bar{\delta}/2, & \text{if } j \in T_y, \\ y_{\tau,j}, & \text{otherwise,} \end{cases}$$

where $\bar{\delta}$ is chosen such that (7) still holds at $(x_\tau, \bar{y}, \bar{t})$.

Obviously, this new point satisfies $(x_\tau, \bar{y}, \bar{t}) \in \mathcal{B}_\delta(x_\tau, y_\tau, t_\tau)$ and so the local optimality of (x_τ, y_τ, t_τ) yields to

$$\begin{aligned}
&\Psi(x_\tau, y_\tau, t_\tau) \leq \Psi(x_\tau, \bar{y}, \bar{t}), \\
&\Theta(x_\tau, y_\tau) + \tau \mathbf{e}^T(t_\tau - y_\tau - h(x_\tau)) + \tau \mathbf{e}^T(t_\tau - g(x_\tau)) \\
&\qquad\qquad \leq \Theta(x_\tau, \bar{y}) + \tau \mathbf{e}^T(\bar{t} - \bar{y} - h(x_\tau)) + \tau \mathbf{e}^T(\bar{t} - g(x_\tau)), \\
&\Theta(x_\tau, y_\tau) - \Theta(x_\tau, \bar{y}) + \tau \mathbf{e}^T(t_\tau - y_\tau) - \tau \mathbf{e}^T(\bar{t} - \bar{y}) + \tau \mathbf{e}^T t_\tau - \tau \mathbf{e}^T \bar{t} \leq 0, \\
&\sum_{j \in T_y} \left(\theta(y_{\tau,j}) - \theta(\bar{y}_j) - \theta(x_{\tau,j} + y_{\tau,j}) + \theta(x_{\tau,j} + \bar{y}_j) + \tau(\bar{y}_j - y_{\tau,j})\right) \\
&\qquad\qquad + 2\tau \sum_{i \in T_t} t_{\tau,i} - \bar{t}_i \leq 0, \\
&\sum_{j \in T_y} \left(\theta(y_{\tau,j}) - \theta(\bar{y}_j) - \theta(x_{\tau,j} + y_{\tau,j}) + \theta(x_{\tau,j} + \bar{y}_j) + \tau(\bar{y}_j - y_{\tau,j})\right) \\
&\qquad\qquad + 2\tau \sum_{i \in T_t} \bar{\delta}/2 \leq 0\ .
\end{aligned}$$

This, however, leads to a contradiction since $\tau, \delta > 0$ and that for all $j \in T_y$

$$\theta(y_{\tau,j}) - \theta(\bar{y}_j) - \theta(x_{\tau,j} + y_{\tau,j}) + \theta(x_{\tau,j} + \bar{y}_j) + \tau(\bar{y}_j - y_{\tau,j}) > 0,$$

since the function $y \mapsto \theta(y) - \theta(x+y) - \tau y$ is non-increasing for all $x, y \geq 0$ and $\tau > \theta'(0)$ by Lemma 4.3. It follows that $T_t = \emptyset$ and $T_y = \emptyset$. Thus, (x_τ, y_τ, t_τ) satisfies (6) and so it is a local minimum of (P). $\square$

5 Difference of Convex Programming

In this section, we introduce some fundamental properties of a DC program and an algorithm to tackle this problem so-called DC Algorithm (DCA). For a complete study on this subject, the readers are referred to [2,23,24] and references herein.

5.1 DC Algorithm

Given two convex functions g, h, a DC program consists in finding $x \in \mathbb{R}^n$ such that

$$\inf_{x \in \mathbb{R}^n} g(x) - h(x). \tag{8}$$

In general this unconstrained formulation encompasses convex constraints by considering $g(x) = g_0(x) + \chi_C(x)$, where g_0 is a convex function, C a closed convex set and $\chi_C(x)$ is the convex indicator function $\chi_C(x) = \{0 \text{ if } x \in C,\ +\infty \text{ if } x \notin C\}$.

Recall that for a lower semi-continuous convex function ψ on $\mathbb{R}^n$ and $x^* \in \mathrm{dom}(\psi) := \{x \mid \psi(x) < \infty\}$, the subdifferential of ψ at x^* is given by

$$\partial\psi(x^*) = \{y \in \mathbb{R}^n \mid \psi(x) \geq \psi(x^*) + (x - x^*)^T y,\ \forall x \in \mathbb{R}^n\}.$$

In the special case, where ψ is a differentiable function, the subdifferential is reduced to a singleton and $\partial\psi(x^*) = \nabla\psi(x^*)$. As an example the subdifferential of the convex indicator function for $x \in C$ is given as

$$\partial\chi_C(x) = \{d \in \mathbb{R}^n \mid d^T(y - x) \leq 0,\ \forall y \in C\} := \mathcal{N}_C(x).$$

The notation $\mathcal{N}_C(x)$ stands for the normal cone of the set C at x.

The necessary local optimality condition for a DC program is

$$\partial h(x^*) \subset \partial g(x^*).$$

We say that x^* satisfies the generalized Kuhn-Tucker condition, if it satisfies

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

We also call such a point a critical point.

DCA is an algorithm that is based on a primal-dual formulation of (8). Let the conjugate function of g be defined as

$$g^*(y) := \sup_{x \in \mathbb{R}^n} x^T y - g(x).$$

Then, the dual of (8) is defined as

$$\inf_{y \in \mathbb{R}^n} h^*(y) - g^*(y). \tag{9}$$

There is no duality gap as proven in [2]. Furthermore, the following result from [2,23,24] gives a characterization of the solutions of both problems. Let P_{dc} (resp. D_{dc}) denotes the set of solutions of the primal DC program (8) (resp. the dual DC program (9)).

Proposition 5.1 *Given g, h two proper lower semi-continuous convex functions. It holds true that*

$$[\cup_{y^* \in D_{dc}} \partial g^*(y^*)] \subset P_{dc} \text{ and } [\cup_{x^* \in P_{dc}} \partial h(x^*)] \subset D_{dc}.$$

Based on local optimality conditions and duality in DCA, the idea of DCA is quite simple: each iteration k of DCA approximates the concave part $-h$ by its affine majorization (that corresponds to take $y^k \in \partial h(x^k)$) and minimizes the resulting convex function (that is equivalent to compute $x^{k+1} \in \partial g^*(y^k)$). The algorithm is stated explicitly in Algorithm 1.

Data: ϵ the precision tolerance; $x^0 \in \mathbb{R}^n$ an initial point;
Begin ;
Set $k := 0$;
repeat
  $y^k \in \partial h(x^k)$;
  $x^{k+1} \in \arg\min\{g(x) - h(x^k) - (x - x^k)^T y^k \ : \ x \in \mathbb{R}^n\}$;
  $k := k + 1$;
until $\|x^k - x^{k-1}\|_\infty \leq \epsilon$;

Algorithm 1: Generic DCA

Convergence properties and its theoretical basis has been analyzed in [2,23, 24]. We remind here the fundamental facts. DCA is a descent method without line search, thus the sequence $\{g(x^k) - h(x^k)\}$ is a decreasing sequence. So, if the optimal value of (8) is finite, then the sequence $\{g(x^k) - h(x^k)\}$ converges. If $g(x^k) - h(x^k) = g(x^{k+1}) - h(x^{k+1})$ or equivalently $x^k = x^{k+1}$ then x^k is a critical point of $g - h$ and the algorithm terminates at iteration k.

Based on the results from Section 4.1 and Section 4.2, DCA can be used to solve the concave NCP and the general NCP. Note here that assuming a solution of the NCP exists, the optimal value is bounded below and the decreasing sequence $\{g(x^k) - h(x^k)\}$ converges.

5.2 DCA for the Concave NCP

We now make the link between DCA and Theorem 3.1. We remind that this theorem proves that a stationary point of (P) with F a concave differentiable P function is the solution of the NCP assuming such a solution exists. The following sequence of results states that under classical assumptions DCA converges to a stationary point of (P).

Lemma 5.1 *Let x^* be a critical point of* (8)*, where we assume that $g(x) := g_0(x) + \chi_C$ with g_0, h differentiable convex functions and C is a closed convex set. It holds true that*

$$0 \in \nabla g_0(x^*) + \mathcal{N}_C(x^*) - \nabla h(x^*),$$

where $\mathcal{N}$ is the normal cone of C at x^*.

Proof x^* is assumed to be a critical point of (8), i.e.

$$0 \in \partial g(x^*) - \partial h(x^*).$$

Now, by construction of g we have

$$0 \in \partial g_0(x^*) + \mathcal{N}_C(x^*) - \partial h(x^*).$$

The result follows by the differentiability assumption on g and h. $\square$

Lemma 5.2 *Let x^* be a critical point of* (8)*, where we assume that $g(x) = g_0(x) + \chi_C$ with g_0, h differentiable lower semi-continuous convex functions and C is a closed convex set. Furthermore, assume that*

$$\mathcal{N}_C(x^*) = \mathscr{L}^\circ(x^*), \tag{10}$$

where

$$\mathscr{L}^\circ(x^*) := \{d \in \mathbb{R}^n \mid d = -\nabla F(x^*)^T\lambda - \mu, \min(F(x^*), \lambda) = 0, \min(x^*, \mu) = 0\}.$$

Then, x^* is stationary point of (P)*.*

The cone $\mathscr{L}^\circ(x^*)$ is sometimes called the polar cone of the linearized cone in the literature related to Karush-Kuhn Tucker conditions, while the condition (10) refers to the Guignard constraint qualification that is known to be the weakest constraint qualification in non-linear programming.

Proof In Theorem 3.1, we define a stationary point of (P) as a point $\bar{x}$ such that there exists Lagrange multipliers $(\lambda, \mu) \in \mathbb{R}^{n \times n}$ that satisfies

$$\begin{aligned}
&\nabla g_0(x^*) - \nabla h(x^*) - \nabla F(x^*)^T\lambda - \mu = 0\\
&\lambda^T F(x^*) = 0, F(x^*) \geq 0, \lambda \geq 0\\
&(x^*)^T\mu = 0, x^* \geq 0, \mu \geq 0\ .
\end{aligned}$$

This can be equivalently written as

$$0 \in \nabla g_0(x^*) - \nabla h(x^*) + \mathscr{L}^\circ(x^*).$$

Now, using assumption (10) we get

$$0 \in \nabla g_0(x^*) - \nabla h(x^*) + \mathcal{N}_C(x^*).$$

However, this condition is satisfied by any critical point x^* according to Lemma 5.1. $\square$

We can now conclude by the following strong result for DC Algorithm applied to (P), which has been shown to be a DC program in Theorem 4.1.

Theorem 5.1 *Let F be a continuously differentiable concave P-function. Assume that a solution of* (1) *exists. Let $\{x^k\}$ be a sequence generated by Algorithm 1 on* (P)*. Then, any limit point x^*, up to a subsequence, of $\{x^k\}$ that satisfies the strict complementarity condition is a solution of* (1)*.*

Proof Assuming that F is a P function and that there exists a solution to (1) yields that this solution is unique. Thus, (P) attains its minimum, whose optimal value is 0.

Let x^* be a limit point of the sequence generated by Algorithm 1.

Applying Lemma 5.2, we get that the limit point x^* is also a stationary point of (P). Indeed, we have proved in Lemma 3.1 that some constraint qualification holds at any feasible point of the problem, which is known to imply (10).

The result follows by applying Theorem 3.1. □

6 Numerics

Through this article, we studied a DC approach to tackle the NCP. In this section, we present four different experiments applying DCA on the DC formulations that have been introduced in the previous sections, formulation (P) for the concave NCP and the penalized formulation (Pen_τ) for the general NCP.

First, we present a comparison on a list of LCPs with other DC approaches that have been suggested recently in [17]. Then, we study an adaptation of our approach to the absolute value equation and present a comparison of existing methods as in [1]. These two families of problems belong to the class of concave NCPs. Finally, we give illustrations of the penalization technique from Section 4.2 on a non-concave NCP and on two examples of large scale NCPs.

The algorithms are coded in *Matlab* on a standard laptop. The convex subproblems of DCA, Algorithm 1, are solved using *CVX* [12]. We consider in all these examples $\theta^1 = \frac{x}{x+1}$.

No attempt has been made to optimize the performance of the algorithm, since our aim is to validate our approach and run a preliminary comparison with other methods. Thus, our main concern is focus on the residuals and in particular on the number of solved problems.

6.1 Comparisons of DC Methods for LCPs

We compare our method denoted TDC with other DC approaches developed in [17] denoted as in the paper DCA1, DCA2, DCA3 and DCA4 with respective initial points $x = 0$ for DCA1 and DCA2, and $x = (0, b)$ for DCA3 and DCA4.

We run our algorithm on a set of 21 problems originally presented in [17]. The precision is set as $\epsilon = 10^{-5}$. The comparative results are given in Table 1, where we reported the objective value of the methods at the final point. The values reported for DCA1, DCA2, DCA3 and DCA4 are those from [17].

Table 1 shows that the method TDC solves all the problem and is then competitive with the best DC approaches from the literature. These experiments already confirm the validity of our approach. The algorithm only required one

Pb.	TDC	DCA1	DCA2	DCA3	DCA4
LCP0	1.01e-08	8.0e-9	1.0e-9	2.56e-6	0.5e-9
LCP1	2.97e-09	4.4e-8	2.0e-9	5.00e-1	1.82e+1
LCP2	6.05e-11	1.2e-9	4.0e-9	7.60e-9	2.3e-9
LCP3	3.51e-09	1.0e-10	1.0e-9	2.21e-4	3.91e+2
LCP4	1.12e-09	1.2e-9	2.0e-9	5.00e-9	3.40e-9
LCP5	7.73e-09	1.3e-9	0.0e-0	1.30e-9	0.0e-0
LCP6	3.44e-11	4.0e-10	1.0e-9	5.98e-2	0.0e-0
LCP6	9.04e-13	6.0e-10	1.0e-9	9.97e-2	0.0e-0
LCP6	1.35e-13	1.0e-10	1.0e-9	12.48e-2	0.0e-0
LCP7	1.48e-10	0.0e-0	3.0e-7	2.92e-6	5.80e-8
LCP7	1.67e-10	1.0e-10	5.0e-7	4.87e-6	7.40e-8
LCP7	1.82e-10	1.0e-10	1.0e-6	9.77e-6	1.20e-7
LCP8	1.03e-10	1.6e-9	3.0e-7	1.14e-6	6.10e-8
LCP8	8.53e-10	2.3e-9	5.0e-7	1.92e-6	1.0e-7
LCP8	5.01e-10	0.0e-0	1.0e-6	3.84e-6	2.00e-7
LCP9	2.24e-09	2.0e-10	1.0e-9	5.98e-2	2.90e-7
LCP9	4.89e-09	2.0e-10	1.0e-9	9.98e-2	6.3e-8
LCP9	4.25e-09	2.0e-10	1.0e-9	1.24e-2	3.30e-8
LCP10	1.45e-09	3.0e-7	0.0e-0	2.0e-10	0.0e-0
LCP10	9.65e-10	5.0e-7	0.0e-0	3.0e-10	0.0e-0
LCP10	9.83e-09	0.0e-0	1.0e-6	6.0e-10	0.0e-0

Table 1 Comparison of the objective values at the solution of five DC methods.

iteration to converge to the solution except for LCP1 (9 iterations) and LCP3 (2 iterations).

6.2 Application to Absolute Value Equations and Comparison

In this section, we apply the same approach to a class of problems that do not belong exactly to the (1) framework but is very close in the sense that it can be reformulated as an NCP with an additional affine equality. This class was widely studied in the last decade and is very often connected to complementarity problems.

In [1], the authors propose a smoothing technique based on a complementarity reformulation of the absolute value equation, defined as

$$Ax - |x| = b, \tag{11}$$

with $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$.

In particular, their motivation was to solve the absolute value equation without any assumption on the data except existence of at least one solution. This is the same motivation as in this article. Using the same technique as in [1], (11) can be cast as the following complementarity problem

$$A(x^+ - x^-) - (x^+ + x^-) = b, \ 0 \leq x^- \perp x^+ \geq 0. \tag{12}$$

Now, applying the technique proposed in this article yields to consider the following sub-additive regularized formulation

$$\begin{aligned} \min_{(x^-,x^+)\in\mathbb{R}^n\times\mathbb{R}^n} \; & \Theta(x^-,x^+) \\ \text{s.t. } & A(x^+-x^-)-(x^++x^-)=b, \\ & (x^-,x^+)\geq 0. \end{aligned} \tag{13}$$

Then, we can solve (11) using the DC approach discussed earlier on (13).

The following heuristic mentioned in [1] can be rather useful to accelerate convergence and assure a good precision when we are close to the solution. After finding the current point x^k we solve, if possible, the linear system

$$(A-\operatorname{diag}(\delta(x^k)))z=b. \tag{14}$$

If x solves (11), then we solved (11) with the same precision as we solved the linear system. However, if x does not solve (11), we continue the iterations with x^k. This idea is similar to compute a Newton iteration.

We generate for several n and several values of the parameters one hundred instances of the problem following [18]:

"Choose a random A from a uniform distribution on $[-10,10]$, then choose a random x from a uniform distribution on $[-1,1]$ and set $b=Ax-|x|$."

The difficulty here relies on the fact that the problem may not be uniquely solvable and thus quite hard to solve. The precision is set as 10^{-6}. We compare 4 methods tailored for general absolute value equations:

- TDC-AVE, which is the DCA on (13) with $\theta(t)=t/(t+0.1)$;
- TAVE method from [1];
- concave minimization method CMM from [18];
- successive linear programming method LPM from [19].

The initial point used for TDC-AVE is obtained by solving the following linear program:

$$\min_{(x^+,x^-)\in\mathbb{R}^n\times\mathbb{R}^n} x^++x^- \text{ s.t. } x^+\geq 0, x^-\geq 0, A(x^+-x^-)-(x^++x^-)=b.$$

This initial point was already proposed in [1].

Due to the potential difficulty of the equation, we focus on the number of unsolved problems. We do not include any Newton-kind method in our comparison, since they are not applicable without additional assumptions on A.

Table 2 shows the number of unsolved problems among 100 problems for each value of n. Table 3 presents the number of iterations and the time required for each method to solve the 100 problems for each value of n.

Overall, these results confirm the validity of our algorithm, since TDC-AVE solves most of the problems in each case (1 failure over 600 instances) and improves significantly the existing methods.

For now, the obvious drawback of this approach is that it is more computationally difficult, since we solve at each step a convex program instead of a

n	TDC-AVE	TAVE	CMM	LPM (LAVE)
32	0	0	9	7
64	0	3	8	13
128	1	8	10	13
256	0	8	11	11
512	0	11	8	10
1024	0	16	7	9

Table 2 Comparison on the number of unsolved problems for 100 hundred randomly generated AVE of size n.

	TDC-AVE	TAVE	CMM	LPM
n	ite.(time)	ite.(time)	ite.(time)	ite.(time)
32	204 (36.92)	306 (0.57)	485 (1.83)	248 (0.79)
64	134 (32.92)	491 (2.76)	458 (2.91)	342 (2.52)
128	207 (108.10)	841 (20.94)	568 (19.05)	409 (17.02)
256	143 (358.70)	1129 (361.38)	595 (182.79)	445 (182.04)
512	127 (2025.1)	1469 (3712.2)	499 (1462.6)	400 (1214.2)
1024	111 (18587)	1993 (37356)	479 (11513)	452 (13565)

Table 3 Comparison on the number of iterations and time in seconds for 100 hundred randomly generated AVE of size n.

linear program. However, it is to be noted that when the dimension grows the gap is reduced. Although, it was not our main goal in this article and we left this for further research.

6.3 A Numerical Example of DC Penalization on a General NCP

In this section, we use the theoretical analysis derived in Section 4.2 to solve the NCP. Let $F : \mathbb{R}^n \to \mathbb{R}^n$.

We proposed in Theorem 4.2 a DC program, which can be interpreted as a penalization of the formulation of the NCP using the sub-additive merit function, (P). The technique used here is to apply DC Algorithm presented in the previous section to the problem (Pen_τ). This approach is only a heuristic, although Theorem 3.1 and Theorem 4.4 tend to give the intuition that it should be a good one.

In order to validate our approach, we report the results applied to an example from [10] and initially proposed in [16], which consider the NCP with

$$F(x) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_2^2 + x_1 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

Two solutions are known: $x^* = (\sqrt{6}/2, 0, 0, 0.5)$ and $\bar{x} = (1, 0, 3, 0)$. The difficulty in solving this problem arises when a Newton-type method is used, since the LCP formed by linearizing F around $x = 0$ has no solution.

Note that all the components of F are convex functions, and so a trivial DC decomposition of F is given by

$$F(x) = g(x) - h(x), \text{ with } g(x) = F(x) \text{ and } h(x) = 0.$$

We choose five random initial points x_0 from a uniform distribution on $[-1, 1]$, we take $\tau_0 = \tau_1 = 1/(r+1), r = 0.1$ and the precision $\epsilon = 10^{-5}$. Results are given in Table 4, where we reported the complementarity $(x^T F(x))$, the solution obtained (sol), the smallest component of F at the solution, the number of iterations (nb-iter), the distance between y and $F(x)$ and the distance between t and $g(x)$.

x^0	$x^T F(x)$	sol	$\|\min(F(x), 0)\|$	nb-iter	$\|y - F(Sol)\|$	$\|t - g(Sol)\|$
x_{01}	0	x^*	0	36	0	0
x_{02}	0	x^*	0	98	0	0
x_{03}	0	$\bar{x}$	0	734	0	0
x_{04}	0	x^*	0	90	0	0
x_{05}	0	x^*	0	61	0	0

Table 4 Sub-additive DC penalization approach to solve the general NCP for five initial points: $x_{01} = (0.4039\ 0.0965\ 0.1320\ 0.9421)^T$, $x_{02} = (0.3532\ 0.8212\ 0.0154\ 0.0430)^T$, $x_{03} = (0.0497\ 0.9027\ 0.9448\ 0.4909)^T$, $x_{04} = (0.4893\ 0.3377\ 0.9001\ 0.3692)^T$ and $x_{05} = (0.6256\ 0.7802\ 0.0811\ 0.9294)^T$.

In these five tests, we have a convergence to one of the two solutions x^* and $\bar{x}$ of the problem. It is interesting to notice that as expected by the theoretical study, when the problem is solved $t = g(x)$ and $y = F(x)$. The results obtained validate the theoretical part and confirm our approach.

6.4 Large-scale NCPs

We conclude these numerical illustrations by reporting some results of the same penalty technique from the previous section applied to larger NCPs.

The two examples P1 and P2 from [14] correspond to strongly monotone function F with

$$F_i(x) = -x_{i+1} + 2x_i - x_{i-1} + \frac{1}{3}x_i^3 - b_i,\ i = 1, \dots, n\ (x_0 = x_{n+1} = 0)$$

and $b_i = (-1)^i$ for the example P1, $b_i = \frac{(-1)^i}{\sqrt{i}}$ for the example P2. In both cases, a DC decomposition of F_i is given by

$$g_i(x) = -x_{i+1} + 2x_i - x_{i-1} + \frac{1}{3}\max(x_i, 0)^3 - b_i \text{ and } h_i(x) = \frac{1}{3}\max(-x_i, 0)^3.$$

We choose the origin as an initial point, we take $\tau_0 = \tau_1 = 1/(r+1), r = 0.1$ and the precision $\epsilon = 10^{-5}$. Results are given in Table 5 for $n \in \{500, 1000, 2000\}$,

P1: n	$x^TF(x)$	$\|\min(x,0)\|_\infty$	$\|\min(F(x),0)\|_\infty$	time
500	4,31e-6	4,62e-11	3,61e-8	18.48
1000	8,90e-6	4,54e-11	3,71e-8	43.88
2000	1,09e-6	5,15e-11	3,14e-9	150.96
P2: n	$x^TF(x)$	$\|\min(x,0)\|_\infty$	$\|\min(F(x),0)\|_\infty$	time
500	9,83e-7	4.16e-10	2,41e-7	21,67
1000	1.21e-6	2.40e-10	1.42e-7	56,57
2000	9,58e-7	4,61e-10	2,26e-7	174,72

Table 5 Sub-additive DC penalization approach to solve the general NCP for Problem P1.

where we reported the complementarity $(x^TF(x))$, the feasibility at the solution and the time in seconds.

In every tests our approach solves the problem up to ϵ, which confirms our approach. As expected the time required to run the algorithm increases with respect to the size.

7 Conclusion

In this paper, we proposed a formulation based on sub-additive merit functions for solving non-linear complementarity problems. We have proved that the reformulated problem is DC program, whenever F is a concave or a DC decomposition of F is known. Thus, we can use the local algorithm DCA to solve the problem.

We also proved that in order to solve the monotone concave NCP, it is sufficient to computing a stationary point of the problem. Besides, when F is a P function and without any additional assumption except that a DC decomposition is known, we propose a penalization formulation, whose local minima are solutions of the NCP.

Numerical experiments on several LCPs and a comparison with other DC approaches prove the validity of our method. We have presented an application of absolute value equation and shows that our method proposed is promising. Finally, we also gave numerical results for some examples when F is non-linear, including large-scale problems, to confirm our penalization approach.

Acknowledgements

The authors would like to thank anonymous referees for their helpful remarks and comments.

References

1. Lina Abdallah, Mounir Haddou, and Tangi Migot. Solving absolute value equation using complementarity and smoothing functions. *Journal of Computational and Applied Mathematics*, 327:196 – 207, 2018.

2. Le Thi Hoai An and Pham Dinh Tao. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of operations research*, 133(1):23–46, 2005.
3. Chunhui Chen and Olvi L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. *Mathematical Programming*, 71(1):51–69, Nov 1995.
4. Chunhui Chen and Olvi L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, 5(2):97–138, Mar 1996.
5. Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The linear complementarity problem.* SIAM, 2009.
6. Tecla De Luca, Francisco Facchinei, and Christian Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, 75(3):407–439, Dec 1996.
7. Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems.* Springer Science & Business Media, 2007.
8. Francisco Facchinei and João Soares. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal on Optimization*, 7(1):225–247, 1997.
9. Andreas Fischer. Solution of monotone complementarity problems with locally lipschitzian functions. *Mathematical Programming*, 76(3):513–532, Mar 1997.
10. Christodoulos A Floudas, Panos M Pardalos, Claire Adjiman, William R Esposito, Zeynep H Gümüs, Stephen T Harding, John L Klepeis, Clifford A Meyer, and Carl A Schweiger. *Handbook of test problems in local and global optimization*, volume 33. Springer Science & Business Media, 2013.
11. Masao Fukushima. *Merit Functions for Variational Inequality and Complementarity Problems*, pages 155–170. Springer US, Boston, MA, 1996.
12. Michael Grant, Stephen Boyd, and Yinyu Ye. Cvx: Matlab software for disciplined convex programming, 2008.
13. Mounir Haddou and Patrick Maheux. Smoothing methods for nonlinear complementarity problems. *Journal of Optimization Theory and Applications*, 160(3):711–729, 2014.
14. Chongchao Huang and Song Wang. A power penalty approach to a Nonlinear Complementarity Problem. *Operation Research Letters*, 38(1):72–76, 2010.
15. Christian Kanzow. Some noninterior continuation methods for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4):851–868, 1996.
16. Masakazu Kojima and Susumu Shindoh. *Extensions of Newton and quasi-Newton methods to systems of PC 1 equations.* Inst. of Technology, Department of Information Sciences, 1986.
17. Hoai An Le Thi and Tao Pham Dinh. On solving linear complementarity problems by dc programming and dca. *Computational Optimization and Applications*, 50(3):507–524, 2011.
18. Olvi L. Mangasarian. Absolute value equation solution via concave minimization. *Optimization Letters*, 1(1):3–8, 2007.
19. Olvi L. Mangasarian. Absolute value equation solution via linear programming. *Journal of Optimization Theory and Applications*, 161(3):870–876, 2014.
20. Tangi Migot and Mounir Haddou. A smoothing method for sparse optimization over polyhedral sets. In *Modelling, computation and optimization in information systems and management sciences*, pages 369–379. Springer, 2015.
21. Mikhail V. Solodov. Constraint qualifications. *Wiley Encyclopedia of Operations Research and Management Science.* Wiley, New York, 2010.
22. Liqun Qi, Defeng Sun, and Guanglu Zhou. A new look at smoothing newton methods for nonlinear complementarity problems and box constrained variational inequalities. *Mathematical Programming*, 87(1):1–35, Jan 2000.
23. Pham Dinh Tao and Le Thi Hoai An. Convex analysis approach to dc programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997.
24. Pham Dinh Tao and Le Thi Hoai An. A dc optimization algorithm for solving the trust-region subproblem. *SIAM Journal on Optimization*, 8(2):476–505, 1998.

25. Pham Dinh Tao and El Bernoussi Souad. *Duality in D.C. (Difference of Convex functions) Optimization. Subgradient Methods*, pages 277–293. Birkhäuser Basel, Basel, 1988.