

How to Compute an M-Stationary Point of the MPCC

Dussault, Jean-Pierre* Haddou, Mounir† Kadrani, Abdeslam‡ Migot, Tangi†

2017

Abstract

We discuss here the convergence of relaxation methods for MPCC with approximate sequence of stationary points by presenting a general framework to study these methods. It has been pointed out in the literature, [23], that relaxation methods with approximate stationary points fail to give guarantee of convergence. We show that by defining a new strong approximate stationarity we can attain the desired goal of computing an M-stationary point. We also provide an algorithmic strategy to compute such point. Existence of strong approximate stationary point in the neighborhood of an M-stationary point is proved.

Keywords : nonlinear programming - MPCC - MPEC - relaxation methods - regularization methods - stationarity - constraint qualification - complementarity - optimization model with complementarity constraints

AMS Subject Classification : 90C30, 90C33, 49M37, 65K05

1 Introduction

We consider the Mathematical Program with Complementarity Constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } g(x) \leq 0, \quad h(x) = 0, \\ 0 \leq G(x) \perp H(x) \geq 0, \end{aligned} \tag{MPCC}$$

with $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $G, H : \mathbb{R}^n \rightarrow \mathbb{R}^q$. All these functions are assumed to be continuously differentiable through this paper. The notation $0 \leq u \perp v \geq 0$ for two vectors u and v in \mathbb{R}^q is a shortcut for $u \geq 0$, $v \geq 0$ and $u^T v = 0$. In this context solving the problem means finding a local minimum. Even so this goal apparently modest is hard to achieve in general.

This problem has become an active subject in the literature since the last two decades and has been the subject of several monographs [26, 27] and PhD thesis [11, 19, 17, 32, 30, 6]. The wide variety of applications that can be casted as an MPCC is one of the reason for this popularity. Among other we can cite truss topology optimization [17], discrete optimization [1], image restoration [3], optimal control [2, 16]. Otherwise, another source of problem are bilevel programming problems [8, 9], where the lower-level problem is replaced by its optimality conditions. This may lead to a more general kind of problem called Mathematical Program with Equilibrium Constraint [27] or Optimization Problem with Variational Inequality Constraint [34]. The MPCC formulation has been the most popular in the literature motivated by more accessible numerical approaches.

(MPCC) is clearly a non-linear programming problem and in general most of the functions involved in the formulation are non-convex. In this context solving the problem means finding a local minimum. Even

*Département d'informatique, Université de Sherbrooke

†IRMAR-Insa

‡INSEA

so this goal apparently modest is hard to achieve in general due to the degenerate nature of the MPCC. Therefore, numerical methods that consider only first order informations may expect to compute a stationary point.

The wide variety of approaches with this aim computes the KKT conditions, which require that some constraint qualification holds at the solution to be an optimality condition. However, it is well-known that these constraint qualifications never hold in general for (MPCC). For instance, the classical Mangasarian-Fromowitz constraint qualification that is very often used to guarantee convergence of algorithms is violated at any feasible point. This is partly due to the geometry of the complementarity constraint that always has an empty relative interior.

These issues have motivated the definition of enhanced constraint qualifications and optimality conditions for (MPCC) as in [34, 33, 28, 13] to cite some of the earliest research. In 2005, Flegel & Kanzow provide an essential result that defines the right necessary optimality condition to (MPCC). This optimality condition is called M(Mordukhovich)-stationary condition. The name comes from the fact that those conditions are derived by using Mordukhovich normal cone in the usual optimality conditions of (MPCC).

In view of the constraint qualifications issues that plague the (MPCC) the relaxation methods provide an intuitive answer. The complementarity constraint is relaxed using a parameter so that the new feasible domain is not thin anymore. It is assumed here that the classical constraints $g(x) \leq 0$ and $h(x) = 0$ are not more difficult to handle than the complementarity constraint. Finally, the relaxing parameter is reduce to converge to the feasible set of (MPCC) similar to an homotopy technique. The interest for such methods has already been the subject of some PhD thesis in [30, 32] and is an active subject in the literature.

These methods have been suggested in the literature back to 2000 by Scheel & Scholtes in [28] replacing the complementarity $\forall i \in \{1, \dots, q\}$ by

$$G_i(x)H_i(x) - t \leq 0.$$

For more clarity we denote $\Phi_t(x)$ the map that relaxed the complementarity constraint and so in this case $\forall i \in \{1, \dots, q\}$

$$\Phi_i^{SS}(G(x), H(x); t) = G_i(x)H_i(x) - t. \quad (\text{SS})$$

This natural approach is later extended by Demiguel, Friedlander, Nogales & Scholtes in [7] by also relaxing the positivity constraints $G(x) \geq -t$, $H(x) \geq -t$. In [25], Lin & Fukushima improve this relaxation by expressing the same set with two constraints instead of three. This improvement leads to improved constraint qualification satisfied by the relaxed sub-problem. Even so the feasible set is not modified this improved regularity does not come as a surprise, since constraint qualification measures the way the feasible set is described and not necessarily the geometry of the feasible set itself. In [31], the authors consider a relaxation of the same type but only around the corner $G(x) = H(x) = 0$ in the following way $\forall i \in \{1, \dots, q\}$

$$\Phi_i^{SU}(G(x), H(x); t) = G_i(x) + H_i(x) - \begin{cases} |G_i(x) - H_i(x)|, & \text{if } |G_i(x) - H_i(x)| \geq t, \\ t\psi\left(\frac{G_i(x) - H_i(x)}{t}\right), & \text{otherwise,} \end{cases} \quad (\text{SU})$$

where ψ is a suitable function as described in [31]. An example of such function being $\psi(z) = \frac{2}{\pi} \sin\left(\frac{\pi}{2}z + \frac{3\pi}{2}\right) + 1$.

In the corresponding papers it has been shown that under suitable conditions providing convergence of the methods they still might converge to some spurious point, called C-stationary point. The convergence to M-stationary being guaranteed only under some second-order condition. Up to this point it is to be noted that different methods used in the literature such as interior-point methods, smoothing of an NCP function and elastic net methods share a lot of common properties with the (SS) method and its extension.

A new perspective for those schemes has been motivated in [20] providing an approximation scheme with convergence to M-stationary point by considering $\forall i \in \{1, \dots, q\}$

$$\Phi_i^{KDB}(G(x), H(x); t) = (G_i(x) - t)(H_i(x) - t). \quad (\text{KDB})$$

This is not a relaxation since the feasible domain of (MPCC) is not included in the feasible set of the subproblems. The method has been extended has a relaxation method through a NCP function in [22]

$\forall i \in \{1, \dots, q\}$ as

$$\Phi_i^{KS}(G(x), H(x); t) = \phi(G_i(x) - t, H_i(x) - t). \quad (\text{KS})$$

In a recent paper, [23], Kanzow & Schwartz discuss convergence of the method (KS) considering sequence of approximate stationary points, that is a point that satisfy approximately the KKT conditions. They illustrate the fact that the method may converge to spurious weak-stationary point. Our motivation in this paper is to deal with this issue and present an algorithmic approach.

We present in this paper a generalized framework to study relaxation methods that encompass the methods KDB, KS and the recent butterfly relaxation. We introduce a new method called asymmetric relaxation that also belong to this framework. Then, by introducing a new kind of approximate stationary point we prove that the methods that belong to the generalized framework converge to M-stationary points. We also deal with the question of existence and computation of such point. Existence is proved in the neighbourhood of an M-stationary point without the need of constraint qualification. An active set-penalization scheme that is a generalization of the penalized-active-set method proposed in [21] is proposed to solve the sub-problems of the relaxation.

The rest of this paper is organized as follows : Section 2 introduces the basic knowledge from non-linear programming and mathematical programming with complementarity constraints that will be extensively used along this paper. Section 3 presents a unified framework to study relaxation methods and Section 4 illustrates that the strongest existing methods as well as a new asymmetric relaxation belong to this framework. Section 5 motivates the difficulty of computing a stationary point of MPCC by using approximate sequences of stationary point. These issues are solved in Section 6, which presents a new notion of approximate stationary point that is sufficient to guarantee the well-behaviour of relaxation methods. Section 7 discuss existence of approximate stationary points in a neighborhood of a solution that may not exist. These issues are handled in two steps. Firstly, Section 8 introduces the MPCC with slack variables. Secondly, Section 9 shows that for the MPCC with slack variables, existence of strong epsilon stationary points is guaranteed under very mild conditions. Finally, Section 10 presents an algorithmic strategy to compute the new approximate stationary point, while Section 11 shows preliminaries numerical results using this strategy.

Notations Through this paper, we use classical notation in optimization. Let x^T denotes the transpose of a vector or a matrix x . The gradient of a function f at a point x with respect to x is denoted $\nabla_x f(x)$ and $\nabla f(x)$ when the derivative is clear from the context. $\text{supp}(x)$ for $x \in \mathbb{R}^n$ is the set of indices such that $x_i \neq 0$ for $i \in \{1, \dots, n\}$. $\mathbb{1}$ is the vector whose components are all one. \mathbb{R}_+ and \mathbb{R}_{++} denotes the set of non-negative and positive real numbers. Given two vectors u and v in \mathbb{R}^n , let $u \circ v$ denotes the Hadamard product of two vectors so that $u \circ v = (u_i v_i)_{1 \leq i \leq n}$. We also use classical asymptotic Landau notations :

$f(x) = o(g(x))$ as $x \rightarrow a$ if and only if for all positive constant M there exists a positive number δ such that $|f(x)| \leq M|g(x)|$ for all $|x - a| \leq \delta$, in other words $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 0$.

$f(x) = \omega(g(x))$ as $x \rightarrow a$ if and only if for all positive constant M there exists a positive number δ such that $|f(x)| \geq M|g(x)|$ for all $|x - a| \leq \delta$.

$f(x) \sim (g(x))$ as $x \rightarrow a$ if and only $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = K$ with K a positive finite constant.

2 Preliminaries

We give classical definitions from non-linear programming and then present their enhanced version to MPCC that will be used in the sequel.

2.1 Non-Linear Programming

Let a general non-linear program be

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } g(x) \leq 0, h(x) = 0, \end{aligned} \quad (\text{NLP})$$

with $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Denote \mathcal{F} the feasible region of (NLP), the set of active indices $\mathcal{I}_g(x) := \{i \in \{1, \dots, p\} \mid g_i(x) = 0\}$ at x , the generalized Lagrangian $\mathcal{L}^r(x, \lambda) = rf(x) + g(x)^T \lambda^g + h(x)^T \lambda^h$ with $\lambda = (\lambda^g, \lambda^h)$ and $\mathcal{M}^r(x)$ is the set of index r multipliers.

By definition, λ is an index r multiplier for (NLP) at a feasible point x if $(r, \lambda) \neq 0$ and $\nabla_x \mathcal{L}^r(x, \lambda) = 0$, $\lambda \geq 0$, $g(x)^T \lambda = 0$. An index 0 multiplier is also called singular multiplier, [4], or an abnormal multiplier, [5]. We call a KKT-point a couple (x, λ) with λ an 1-index multiplier at x . A couple (x, λ) with λ a 0-index multiplier at x is called Fritz-John point.

In the context of solving non-linear program, that is finding a local minimum, one widely used technique is to compute necessary conditions also called optimality conditions. The principle tool is the Karush-Kuhn-Tucker (KKT) conditions. Let x^* be a local minimum of (NLP) that satisfy a constraint qualification, then there exists $\lambda \in \mathcal{M}^1(x^*)$ that satisfy

$$\begin{aligned} \nabla_x \mathcal{L}^1(x^*, \lambda) &= 0, \\ \min(-g(x^*), \lambda^g) &= 0, h(x^*) = 0. \end{aligned} \quad (\text{KKT})$$

A point (x^*, λ) that satisfy (KKT) is called a stationary point.

In the context of numerical computation it can be difficult to compute stationary points. Hence, it is of interest to consider epsilon-stationary points.

Definition 2.1 (epsilon-stationary point). *Given a general non-linear program (NLP) and $\epsilon \geq 0$. We say that $(x^*, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is an epsilon-stationary point (or an epsilon-KKT point) if it satisfies*

$$\|\nabla_x \mathcal{L}^1(x^*, \lambda)\|_\infty \leq \epsilon,$$

and

$$\begin{aligned} g_i(x^*) &\leq \epsilon, \lambda_i^g \geq 0, |\lambda_i^g g_i(x^*)| \leq \epsilon, \forall i \in \{1, \dots, p\}, \\ |h_i(x^*)| &\leq \epsilon, \forall i \in \{1, \dots, m\}. \end{aligned}$$

At $\epsilon = 0$ we get the classical definition of a stationary point of (NLP).

2.2 Mathematical Program with Complementarity Constraints

Let \mathcal{Z} be the set of feasible points of (MPCC). Given $x \in \mathcal{Z}$, we denote

$$\begin{aligned} \mathcal{I}^{+0}(x) &:= \{i \in \{1, \dots, q\} \mid G_i(x) > 0 \text{ and } H_i(x) = 0\}, \\ \mathcal{I}^{0+}(x) &:= \{i \in \{1, \dots, q\} \mid G_i(x) = 0 \text{ and } H_i(x) > 0\}, \\ \mathcal{I}^{00}(x) &:= \{i \in \{1, \dots, q\} \mid G_i(x) = 0 \text{ and } H_i(x) = 0\}. \end{aligned}$$

We define the generalized MPCC-Lagrangian function of (MPCC) as

$$\mathcal{L}_{MPCC}^r(x, \lambda) := rf(x) + \lambda^g g(x) + \lambda^h h(x) - \lambda^G G(x) - \lambda^H H(x),$$

with $\lambda := (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$.

We remind that the tangent cone of a set X at $x^* \in X$ is a closed cone defined by

$$\mathcal{T}_X(x^*) := \{d \in \mathbb{R}^n \mid \exists t_k \geq 0 \text{ and } x^k \rightarrow_X x^* \text{ s.t. } t_k(x^k - x^*) \rightarrow d\}.$$

The cone \mathcal{L}_{MPCC} , defined in [28], is given as the following not necessarily convex cone

$$\begin{aligned} \mathcal{L}_{MPCC}(x^*) := \{d \in \mathbb{R}^n \mid & \nabla g_i(x^*)^T d \leq 0 \ \forall i = 1, \dots, p, \\ & \nabla h_i(x^*)^T d = 0 \ \forall i = 1, \dots, m, \\ & \nabla G_i(x^*)^T d = 0 \ \forall i \in \mathcal{I}^{0+}(x^*), \\ & \nabla H_i(x^*)^T d = 0 \ \forall i \in \mathcal{I}^{+0}(x^*), \\ & \nabla G_i(x^*)^T d \geq 0, \nabla H_i(x^*)^T d \geq 0 \ \forall i \in \mathcal{I}^{00}(x^*), \\ & (\nabla G_i(x^*)^T d)(\nabla H_i(x^*)^T d) = 0 \ \forall i \in \mathcal{I}^{00}(x^*)\}. \end{aligned}$$

Due to [12], one always has the following inclusions

$$\mathcal{T}_{\mathcal{Z}}(x^*) \subseteq \mathcal{L}_{MPCC}(x^*).$$

Given a cone $K \subset \mathbb{R}^n$, the polar of K is the cone defined by $K^\circ := \{z \in \mathbb{R}^n \mid z^T x \leq 0, \forall x \in K\}$.

We can now define a mild constraint qualification for (MPCC) called MPCC-Guignard CQ.

Definition 2.2. *Let $x^* \in \mathcal{Z}$. We say that MPCC-GCQ holds at x^* if $\mathcal{T}_{\mathcal{Z}}^\circ(x^*) = \mathcal{L}_{MPCC}^\circ(x^*)$.*

In general, there does not exist KKT stationary points since (MPCC) is highly degenerate and does not satisfy classical constraint qualification from non-linear programming. So we introduce weaker stationary concepts as in [28, 33].

Definition 2.3. *[Stationary point] $x^* \in \mathcal{Z}$ is said*

- *Weak-stationary if there exists $\lambda \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$ such that*

$$\begin{aligned} \nabla_x \mathcal{L}_{MPCC}^1(x^*, \lambda) &= 0, \\ \min(-g(x^*), \lambda^g) &= 0, \quad h(x^*) = 0, \\ \forall i \in \mathcal{I}^{+0}(x^*), \lambda_i^G &= 0, \quad \text{and } \forall i \in \mathcal{I}^{0+}(x^*), \lambda_i^H = 0; \end{aligned}$$

- *Clarke-stationary point if x^* is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \lambda_i^G \lambda_i^H \geq 0;$$

- *Alternatively (or Abadie)-stationary point if x^* is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \lambda_i^G \geq 0 \text{ or } \lambda_i^H \geq 0;$$

- *Mordukhovich-stationary point if x^* is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \text{ either } \lambda_i^G > 0, \lambda_i^H > 0 \text{ or } \lambda_i^G \lambda_i^H = 0;$$

- *Strong-stationary point if x^* is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \lambda_i^G \geq 0, \lambda_i^H \geq 0.$$

Relations between these notions are straightforward from the definition.

The following theorem is a keystone to define necessary optimality conditions for (MPCC).

Theorem 2.1 ([14]). *A local minimum of (MPCC) that satisfies MPCC-GCQ or any stronger MPCC-CQ is an M-stationary point.*

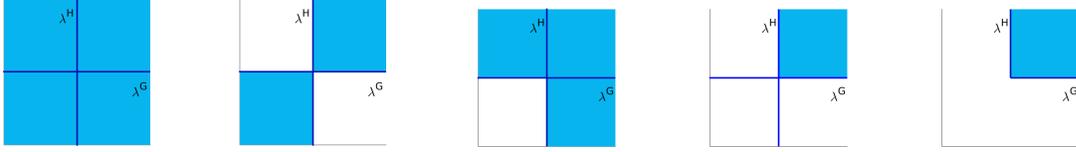


Figure 1: Signs of λ^G, λ^H for indices $i \in \mathcal{I}^{00}$. From the left to the right : weak-stationary, C-stationary, A-stationary, M-stationary and S-stationary.

Therefore, devising algorithms to reach KKT stationary points (S-stationary) is not possible in general, and we must satisfy ourselves in devising algorithms reaching M-stationary points. The following example due to Kanzow and Schwartz exhibits a situation where the global minimizer is not a KKT point but an M-stationary point. We will return to this example later on.

Example 2.1.

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & x_1 + x_2 - x_3 \\ \text{s.t.} \quad & g_1(x) := -4x_1 + x_3 \leq 0, \\ & g_2(x) := -4x_2 + x_3 \leq 0, \\ & 0 \leq G(x) := x_1 \perp H(x) := x_2 \geq 0. \end{aligned}$$

The global solution is $(0, 0, 0)^t$ but is not a KKT point. Indeed, the gradient of the Lagrangian equal to zero yields

$$0 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \lambda_1^g \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + \lambda_2^g \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} - \lambda_1^G \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \lambda_2^H \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \eta \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

and since $\lambda_1^g + \lambda_2^g = 1$ (third line), summing the first two lines yields $2 - 4(\lambda_1^g + \lambda_2^g) - \lambda_1^G - \lambda_2^H = 0$ and therefore $\lambda_1^G + \lambda_2^H = -2$; both cannot be non-negative.

Apart from MPCC-GCQ there exists a wide variety of MPCC constraint qualification described in the literature. We conclude this section by defining only a short selection of them with interesting properties for relaxation methods. Most of these conditions are constraint qualification from non-linear programming that are extended for (MPCC).

One of the most principal constraint qualification used in the literature of (MPCC) is the MPCC-LICQ, see [29] for a discussion on this CQ. In a similar way we extend CRCQ as in [15]. A condition that is similar was used in [22, 18] to prove convergence of relaxation methods for (MPCC).

Definition 2.4. Let $x^* \in \mathcal{Z}$.

1. MPCC-LICQ holds at x^* if the gradients

$$\{\nabla g_i(x^*) (i \in \mathcal{I}_g(x^*)), \nabla h_i(x^*) (i = 1, \dots, m), \nabla G_i(x^*) (i \in \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*)), \nabla H_i(x^*) (i \in \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{+0}(x^*))\}$$

are linearly independent.

2. MPCC-CRCQ holds at x^* if there exists $\delta > 0$ such that, for any subsets $\mathcal{I}_1 \subseteq \mathcal{I}_g(x^*)$, $\mathcal{I}_2 \subseteq \{1, \dots, m\}$, $\mathcal{I}_3 \subseteq \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*)$, and $\mathcal{I}_4 \subseteq \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*)$, the family of gradients

$$\{\nabla g_i(x^*) (i \in \mathcal{I}_1), \nabla h_i(x^*) (i \in \mathcal{I}_2), \nabla G_i(x^*) (i \in \mathcal{I}_3), \nabla H_i(x^*) (i \in \mathcal{I}_4)\}$$

has the same rank for each $x \in \mathbb{B}_\delta(x^*)$, where $\mathbb{B}_\delta(x^*)$ is the ball of radius δ centered at x^* .

In order to prove convergence of very general relaxation methods we consider the definition of MPCC-CRSC, which was introduced and proved to be a constraint qualification very recently in [10].

The polar of the cone \mathcal{L}_{MPCC} is a key tool in the definition of constraint qualification. It is however not trivial to compute. Therefore, we introduce the following cone:

$$\begin{aligned} \mathcal{P}_M(x^*) &:= \{d \in \mathbb{R}^n \mid \exists(\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in \mathbb{R}_+^p \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q \\ &\quad \text{with } \lambda_i^G \lambda_i^H = 0 \text{ or } \lambda_i^G > 0, \lambda_i^H > 0 \forall i \in \mathcal{I}^{00}(x^*), \\ d &= \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^m \lambda_i^h \nabla h_i(x^*) \\ &\quad - \sum_{i \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda_i^G \nabla G_i(x^*) - \sum_{i \in \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda_i^H \nabla H_i(x^*) \}. \end{aligned}$$

Definition 2.5. *Let $x^* \in \mathcal{Z}$. MPCC-CRSC holds at x^* if for any partition $\mathcal{I}_{++}^{00} \cup \mathcal{I}_{0-}^{00} \cup \mathcal{I}_{-0}^{00} = \mathcal{I}^{00}(x^*)$ such that*

$$\begin{aligned} &\sum_{i \in \mathcal{I}_g} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^m \lambda_i^h \nabla h_i(x^*) - \sum_{i \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}_{++}^{00}} \lambda_i^G \nabla G_i(x^*) - \sum_{i \in \mathcal{I}^{+0}(x^*) \cup \mathcal{I}_{++}^{00}} \lambda_i^H \nabla H_i(x^*) \\ &+ \sum_{i \in \mathcal{I}_{-0}^{00}} \lambda_i^G \nabla G_i(x^*) + \sum_{i \in \mathcal{I}_{0-}^{00}} \lambda_i^H \nabla H_i(x^*) = 0, \end{aligned}$$

with $\lambda_i^g \geq 0$ ($i \in \mathcal{I}_g(x^*)$), λ_i^G and $\lambda_i^H \geq 0$ ($i \in \mathcal{I}_{++}^{00}$), $\lambda_i^G > 0$ ($i \in \mathcal{I}_{-0}^{00}$), λ_i^H ($i \in \mathcal{I}_{0-}^{00}$) > 0 , there exists $\delta > 0$ such that the family of gradients

$$\{\nabla g_i(x) \ (i \in \mathcal{I}_1), \ \nabla h_i(x) \ (i = 1, \dots, m), \ \nabla G_i(x) \ (i \in \mathcal{I}_3), \ \nabla H_i(x) \ (i \in \mathcal{I}_4)\},$$

has the same rank for every $x \in \mathcal{B}_\delta(x^*)$, where

$$\begin{aligned} \mathcal{I}_1 &:= \{i \in \mathcal{I}_g(x^*) \mid -\nabla g_i(x^*) \in \mathcal{P}_M(x^*)\}, \\ \mathcal{I}_3 &:= \mathcal{I}^{0+}(x^*) \cup \{i \in \mathcal{I}_{++}^{00} \mid \nabla G_i(x^*) \in \mathcal{P}_M(x^*)\} \cup \mathcal{I}_{-0}^{00}, \\ \mathcal{I}_4 &:= \mathcal{I}^{+0}(x^*) \cup \{i \in \mathcal{I}_{++}^{00} \mid \nabla H_i(x^*) \in \mathcal{P}_M(x^*)\} \cup \mathcal{I}_{0-}^{00}. \end{aligned}$$

It is not necessary to add that the gradients $-\nabla G_i(x^*)$ and $-\nabla H_i(x^*)$ belong to $\mathcal{P}_M(x^*)$. Indeed, we already require that λ_i^G and λ_i^H must be non-zero respectively for the indices $i \in \mathcal{I}_{-0}^{00}$ and $i \in \mathcal{I}_{0-}^{00}$ and so it implies that these gradients belong to this set.

Furthermore, MPCC-CRSC is weaker than MPCC-CRCQ. Indeed, MPCC-CRCQ requires that every family of linearly dependant gradients remains linearly dependant in some neighbourhood, while the MPCC-CRSC condition consider only the family of gradients that are linearly dependant with coefficients that have M-stationary signs.

3 A Unified Framework for Relaxation/Approximation Methods

In the past decade, several methods have been proposed to compute an M-stationary point of (MPCC). The first was the approximation scheme proposed by [20], which was latter improved as a relaxation by [22]. This relaxation scheme has been generalized recently in [10] to a more general family of relaxation schemes. We proposed in this section a unified framework that embraces those methods and may be used to derive new ones.

Consider the following parametric non-linear program $R_t(x)$ parametrized by the vector t :

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{s.t. } g(x) \leq 0, \quad h(x) = 0, \\ & \quad G(x) \geq -\bar{t}(t)\mathbb{1}, \quad H(x) \geq -\bar{t}(t)\mathbb{1}, \quad \Phi(G(x), H(x); t) \leq 0, \end{aligned} \tag{R_t(x)}$$

with $\bar{t} : \mathbb{R}_+^l \rightarrow \mathbb{R}_+$ such that $\lim_{\|t\| \rightarrow 0} \bar{t}(t) \rightarrow 0$ and the relaxation map $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^q$. In the sequel we skip the dependency in t and denote \bar{t} to simplify the notation. It is to be noted here that t is a vector of an arbitrary size denoted l as for instance in [10] where $l = 2$. The generalized Lagrangian function of $(R_t(x))$ is defined for $\nu \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^q$ as

$$\mathcal{L}_{R_t}^r(x, \nu) := rf(x) + g(x)^T \nu^g + h(x)^T \nu^h - G(x)^T \nu^G - H(x)^T \nu^H + \Phi(G(x), H(x); t)^T \nu^\Phi.$$

Let \mathcal{I}_Φ be the set of active indices for the constraint $\Phi(G(x), H(x); t) \leq 0$, i.e.

$$\mathcal{I}_\Phi(x; t) := \{i \in \{1, \dots, q\} \mid \Phi_i(G(x), H(x); t) = 0\}.$$

The definition of a generic relaxation scheme is completed by the following hypotheses:

- $\Phi(G(x), H(x); t)$ is a continuously differentiable real valued map extended component by component, so that

$$\Phi_i(G(x), H(x); t) := \Phi(G_i(x), H_i(x); t). \tag{H1}$$

- Direct computations give that the gradient with respect to x for $i \in \{1, \dots, q\}$ of $\Phi_i(G(x), H(x); t)$ for all $x \in \mathbb{R}^n$ is given by

$$\nabla_x \Phi_i(G(x), H(x); t) = \nabla G_i(x) \alpha_i^H(x; t) + \nabla H_i(x) \alpha_i^G(x; t),$$

where $\alpha^H(x; t)$ and $\alpha^G(x; t)$ are continuous maps by smoothness assumption on $\Phi(G(x), H(x); t)$, which we assume satisfy $\forall x \in \mathcal{Z}$

$$\lim_{\|t\| \rightarrow 0} \alpha^H(x; t) = H(x) \quad \text{and} \quad \lim_{\|t\| \rightarrow 0} \alpha^G(x; t) = G(x). \tag{H2}$$

- At the limit when $\|t\|$ goes to 0, the feasible set of the parametric non-linear program $(R_t(x))$ must converges to the feasible set of (MPCC). In other words, given $\mathcal{F}(t)$ the feasible set of $(R_t(x))$ it holds that

$$\lim_{\|t\| \rightarrow 0} \mathcal{F}(t) = \mathcal{Z}, \tag{H3}$$

where the limit is assumed pointwise.

- At the boundary of the feasible set of the relaxation of the complementarity constraint it holds that for all $i \in \{1, \dots, q\}$

$$\Phi_i(G(x), H(x); t) = 0 \implies F_{G_i}(x; t) = 0 \quad \text{or} \quad F_{H_i}(x; t) = 0, \tag{H4}$$

where

$$\begin{aligned} F_G(x; t) &:= G(x) - \psi(H(x); t), \\ F_H(x; t) &:= H(x) - \psi(G(x); t), \end{aligned} \tag{1}$$

and ψ is a continuously differentiable real valued function extended component by component. Note that the function ψ may be two different functions in (1) as long as they satisfy the assumptions below. Those functions $\psi(H(x); t)$, $\psi(G(x); t)$ are non-negative for all $x \in \{x \in \mathbb{R}^n \mid \Phi(G(x), H(x); t) = 0\}$ and satisfy $\forall z \in \mathbb{R}^q$

$$\lim_{\|t\| \rightarrow 0} \psi(z; t) = 0. \tag{H5}$$

We will prove in Lemma 6.1 that this generic relaxation scheme for (MPCC) converges to an M-stationary point requiring the following essential assumption on the functions ψ . As t goes to 0 the derivative with respect to the first variable of ψ satisfies $\forall z \in \mathbb{R}^q$

$$\lim_{\|t\| \rightarrow 0} \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=z} = 0. \quad (\text{H6})$$

We conclude this section by giving an explicit formula for the relaxation map at the boundary of the feasible set.

Lemma 3.1. *Given $\Phi(G(x), H(x); t)$ be such that for all $i \in \mathcal{I}_\Phi(x; t)$*

$$\Phi_i(G(x), H(x); t) = F_{G_i}(x; t) F_{H_i}(x; t).$$

The gradient with respect to x of $\Phi_i(G(x), H(x); t)$ for $i \in \mathcal{I}_\Phi(x; t)$ is given by

$$\nabla_x \Phi_i(G(x), H(x); t) := \nabla G_i(x) \alpha_i^H(x; t) + \nabla H_i(x) \alpha_i^G(x; t),$$

with

$$\begin{aligned} \alpha_i^G(x; t) &= F_{G_i}(x; t) - \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=H_i(x)} F_{H_i}(x; t), \\ \alpha_i^H(x; t) &= F_{H_i}(x; t) - \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=G_i(x)} F_{G_i}(x; t). \end{aligned}$$

4 Existing Methods under the Unified Framework

In this section, we illustrate the fact that the existing methods in the literature fall under this unified framework. Indeed, the approximation method from Kadrani et al. [20] as well as the two relaxation methods from Kanzow & Schwartz [22] and from Dussault, Haddou & Migot [10] satisfy those hypothesis. We conclude this section by presenting a new asymmetric relaxation method that also belong to our framework.

An optimization method that satisfies all of the 6 hypothesis defined in the previous section is called an UF-method.

4.1 The Boxes Approximation

In 2009 Kadrani, Dussault and Bechakroun introduce a method, which enjoys the desired goal to converge to an M-stationary point, see [20]. Their original method consider an approximation of the complementarity constraints as a union of two boxes connected only on one point (t, t) (here $t \in \mathbb{R}$), in the following way $\forall i \in \{1, \dots, q\}$:

$$\Phi_i^{KDB}(x; t) = (G_i(x) - t)(H_i(x) - t). \quad (2)$$

This is not a relaxation but an approximation, since the feasible domain of the relaxed problem does not include the feasible domain of (MPCC) as illustrated on Figure 2.

Proposition 4.1. *The approximation scheme $(R_t(x))$ with (2) is an UF-method.*

Proof. Continuity of the map Φ as well as (H3) has been proved in [20].

(H4) is satisfied by construction considering $\psi(z; t) = t$. In this case (H6) and (H5) are obviously satisfied.

Now, we consider (H2). Direct computations give that the gradient of Φ for all $i \in \{1, \dots, q\}$ is given by

$$\nabla_x \Phi_i^{KDB}(x; t) = \nabla G_i(x)(H_i(x) - t) + \nabla H_i(x)(G_i(x) - t).$$

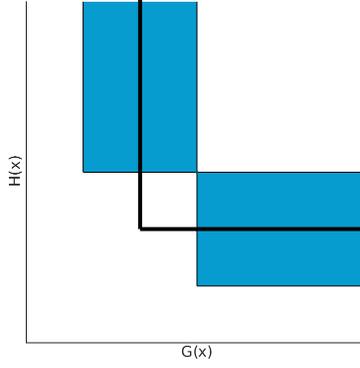


Figure 2: Feasible set of the approximation (2).

Therefore, α_i^G and α_i^H are given by

$$\begin{aligned}\alpha_i^H(x; t) &= H_i(x) - t, \\ \alpha_i^G(x; t) &= G_i(x) - t.\end{aligned}$$

It clearly holds that $\alpha_i^G(x; t) \rightarrow_{\|t\| \rightarrow 0} G_i(x)$ and $\alpha_i^H(x; t) \rightarrow_{\|t\| \rightarrow 0} H_i(x)$. So, in this case (H2) is satisfied.

This completes the proof that all of the 6 hypothesis are satisfied and so the approximation (2) is an UF-method. \square

4.2 The L-shape Relaxation

The previous method has latter been extended to a relaxation in [22] as illustrated on Figure 3 using a piecewise NCP function by considering $\forall i \in \{1, \dots, q\}$

$$\Phi_i^{KS}(x; t) = \phi(G_i(x) - t, H_i(x) - t), \quad (3)$$

where $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a continuously differentiable NCP-function with for instance

$$\phi(a, b) = \begin{cases} ab, & \text{if } a + b \geq 0, \\ -\frac{1}{2}(a^2 + b^2), & \text{if } a + b < 0. \end{cases}$$

Proposition 4.2. *The relaxation scheme $(R_t(x))$ with (3) is an UF-method.*

Proof. Continuity of the map Φ as well as (H3) has been proved in [22].

(H4) is satisfied by construction considering $\psi(z; t) = t$. In this case (H6) and (H5) are obviously satisfied.

Now, we consider (H2). Direct computations give that the gradient of Φ for all $i \in \{1, \dots, q\}$ is given by

$$\nabla_x \Phi_i^{KS}(x; t) = \begin{cases} \nabla G_i(x)(H_i(x) - t) + \nabla H_i(x)(G_i(x) - t), & \text{if } H_i(x) - t + G_i(x) - t \geq 0, \\ -\nabla G_i(x)(G_i(x) - t) - \nabla H_i(x)(H_i(x) - t), & \text{else.} \end{cases}$$

Therefore, α_i^G and α_i^H are given by

$$\begin{aligned}\alpha_i^H(x; t) &= \begin{cases} H_i(x) - t, & \text{if } H_i(x) - t + G_i(x) - t \geq 0, \\ -(G_i(x) - t), & \text{otherwise,} \end{cases} \\ \alpha_i^G(x; t) &= \begin{cases} G_i(x) - t, & \text{if } H_i(x) - t + G_i(x) - t \geq 0, \\ -(H_i(x) - t), & \text{otherwise.} \end{cases}\end{aligned}$$

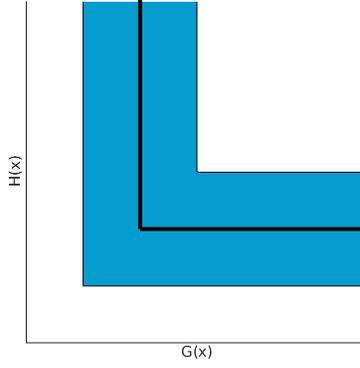


Figure 3: Feasible set of the relaxation (3).

In the case $H_i(x) - t + G_i(x) - t \geq 0$ it clearly holds that $\alpha_i^G(x; t) \rightarrow G_i(x)$ and $\alpha_i^H(x; t) \rightarrow H_i(x)$. So, in this case (H2) is satisfied.

In the case $H_i(x) - t + G_i(x) - t < 0$ the opposite holds that is $\alpha_i^G(x; t) \rightarrow -H_i(x)$ and $\alpha_i^H(x; t) \rightarrow -G_i(x)$. However, it is to be noted that sequences x^t with $x^t \rightarrow_{\|t\| \rightarrow 0} x^*$ that belong to this case satisfy $i \in \mathcal{I}^{00}(x^*)$. To sum up, in this case for $x \in \mathcal{Z}$ then $\alpha_i^G(x; t) \rightarrow H_i(x) = G_i(x) = 0$ and $\alpha_i^H(x; t) \rightarrow G_i(x) = H_i(x) = 0$. This proves that (H2) holds in this case too and so this hypothesis holds for this relaxation.

This completes the proof that all of the 6 hypothesis are satisfied and so the relaxation (3) is an UF-method. \square

4.3 The Butterfly Relaxation

The butterfly family of relaxations deal with two positive parameters (t_1, t_2) defined such that for all $i \in \{1, \dots, q\}$

$$\Phi_i^B(x; t) := \phi(F_{1i}(x; t), F_{2i}(x; t)), \quad (4)$$

with

$$\begin{aligned} F_{1i}(x; t) &:= H_i(x) - t_1 \theta_{t_2}(G_i(x)), \\ F_{2i}(x; t) &:= G_i(x) - t_1 \theta_{t_2}(H_i(x)), \end{aligned}$$

where $\theta_{t_2} : \mathbb{R} \rightarrow]-\infty, 1]$ are continuously differentiable non-decreasing concave function with $\theta(0) = 0$, and $\lim_{t_2 \rightarrow 0} \theta_{t_2}(x) = 1 \forall x \in \mathbb{R}_{++}$ completed in a smooth way for negative values by considering $\theta_{t_2}(z < 0) = \frac{z\theta'(0)t_1}{t_2}$.

We assume the following relation between the parameter

$$t_1 = o(t_2) \text{ and } t_1 = \omega(t_2^2).$$

This method is illustrated on Figure 4.

Proposition 4.3. *The relaxation scheme $(R_t(x))$ with (4) is an UF-method.*

Proof. Continuity of the map Φ as well as (H3) has been proved in [10].

(H4) is satisfied by construction considering $\psi(z; t) = t_1 \theta_{t_2}(z)$. In this case (H5) and (H6) are obviously satisfied. The latter being insured by $t_1 = o(t_2)$.

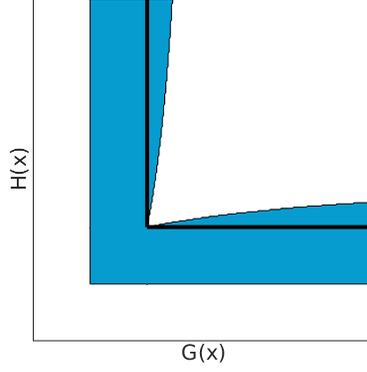


Figure 4: Feasible set of the relaxation (4).

Now, we consider (H2). Direct computations give that the gradient of Φ for all $i \in \{1, \dots, q\}$ is given by

$$\nabla_x \Phi_i^B(x; t) = \begin{cases} (F_{1_i}(x; t) - t_1 \theta'_{t_2}(G_i(x)) F_{2_i}(x; t)) \nabla G_i(x) \\ + (F_{2_i}(x; t) - t_1 \theta'_{t_2}(H_i(x)) F_{1_i}(x; t)) \nabla H_i(x), & \text{if } F_{1_i}(x; t) + F_{2_i}(x; t) \geq 0, \\ (t_1 \theta'_{t_2}(G_i(x)) F_{1_i}(x; t) - F_{2_i}(x; t)) \nabla G_i(x) \\ + (t_1 \theta'_{t_2}(H_i(x)) F_{2_i}(x; t) - F_{1_i}(x; t)) \nabla H_i(x), & \text{if } F_{1_i}(x; t) + F_{2_i}(x; t) < 0. \end{cases}$$

Therefore, α_i^G and α_i^H are given by

$$\alpha_i^H(x; t) = \begin{cases} F_{1_i}(x; t) - t_1 \theta'_{t_2}(G_i(x)) F_{2_i}(x; t), & \text{if } F_{1_i}(x; t) + F_{2_i}(x; t) \geq 0, \\ t_1 \theta'_{t_2}(G_i(x)) F_{1_i}(x; t) - F_{2_i}(x; t), & \text{otherwise,} \end{cases}$$

$$\alpha_i^G(x; t) = \begin{cases} F_{2_i}(x; t) - t_1 \theta'_{t_2}(H_i(x)) F_{1_i}(x; t), & \text{if } F_{1_i}(x; t) + F_{2_i}(x; t) \geq 0, \\ t_1 \theta'_{t_2}(H_i(x)) F_{2_i}(x; t) - F_{1_i}(x; t), & \text{otherwise.} \end{cases}$$

In the case $F_{1_i}(x; t) + F_{2_i}(x; t) \geq 0$ it clearly holds that $\alpha_i^G(x; t) \rightarrow G_i(x)$ and $\alpha_i^H(x; t) \rightarrow H_i(x)$. So, in this case (H2) is satisfied.

In the case $F_{1_i}(x; t) + F_{2_i}(x; t) < 0$ the opposite holds that is $\alpha_i^G(x; t) \rightarrow -H_i(x)$ and $\alpha_i^H(x; t) \rightarrow -G_i(x)$. However, it is to be noted that sequences x^t with $x^t \rightarrow_{\|t\| \rightarrow 0} x^*$ that belongs to this case satisfy $i \in \mathcal{I}^{00}(x^*)$. Therefore, in this case for $x \in \mathcal{Z}$ then $\alpha_i^G(x; t) \rightarrow H_i(x) = G_i(x) = 0$ and $\alpha_i^H(x; t) \rightarrow G_i(x) = H_i(x) = 0$. This proves that (H2) holds in this case too and so this hypothesis holds for this relaxation.

This completes the proof that all of the 6 hypothesis are satisfied and so the relaxation (4) is an UF-method. \square

4.4 An Asymmetric Relaxation

Up till now we only consider relaxation methods that are symmetric. We can define also asymmetric relaxation methods illustrated on Figure 5 that respect the hypothesis of our unified framework.

Let I_G and I_H be two sets of indices such that $I_G \cup I_H = \{1, \dots, q\}$ and $I_G \cap I_H = \emptyset$. Then, the relaxation constraint is defined with

$$\Phi_i(G(x), H(x); t) = \begin{cases} (G_i(x) - t)H_i(x), & \text{for } i \in I_G, \\ G_i(x)(H_i(x) - t), & \text{for } i \in I_H. \end{cases} \quad (5)$$

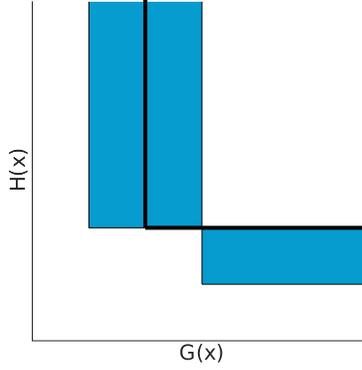


Figure 5: Feasible set of the relaxation (5) with $I_H = \{1\}$ and $I_G = \emptyset$.

Proposition 4.4. *The relaxation scheme $(R_t(x))$ with (5) is an UF-method.*

Proof. Continuity of the map $\Phi(G(x), H(x); t)$ as well as (H3) can be easily deduce from the definition of (5).

(H4) is satisfied by construction considering $\psi(z; t) = t$ or 0. In this case (H5) and (H6) are obviously satisfied.

Now, we consider (H2). Direct computations give that the gradient of Φ for all $i \in \{1, \dots, q\}$ is given by

$$\nabla_x \Phi_i(G(x), H(x); t)(x) = \begin{cases} \nabla G_i(x) H_i(x) + \nabla H_i(x) (G_i(x) - t), & \text{for } i \in I_G, \\ \nabla G_i(x) (H_i(x) - t) + \nabla H_i(x) G_i(x), & \text{for } i \in I_H. \end{cases}$$

Therefore, α_i^G and α_i^H are given by

$$\alpha_i^H(x; t) = \begin{cases} H_i(x), & \text{for } i \in I_G, \\ H_i(x) - t, & \text{for } i \in I_H, \end{cases}$$

$$\alpha_i^G(x; t) = \begin{cases} G_i(x) - t, & \text{for } i \in I_G, \\ G_i(x), & \text{for } i \in I_H. \end{cases}$$

Clearly in both cases (H2) is satisfied.

This completes the proof that all of the 6 hypothesis are satisfied and so the relaxation (3) is an UF-method. \square

5 Motivations on Epsilon-Solution to the Regularized Subproblems

We have seen in the previous sections a general framework to define relaxations of (MPCC). From an algorithmic point of view, the main idea of relaxation methods to solve (MPCC) is to compute a sequence of stationary points or more precisely approximate stationary points for each value of a sequence of parameter $\{t_k\}$. The following definition is a specialization of Definition 2.1 for $(R_t(x))$. It consists in replacing most “0” in (KKT) by small quantities ϵ .

Definition 5.1. x^k is an epsilon-stationary point for $(R_t(x))$ with $\epsilon_k \geq 0$ if there exists $\nu^k \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^q$ such that

$$\|\nabla \mathcal{L}_{R_t}^1(x^k, \nu^k; t_k)\|_\infty \leq \epsilon_k$$

and

$$\begin{aligned}
g_i(x^k) &\leq \epsilon_k, \nu_i^{g,k} \geq 0, |g_i(x^k)\nu_i^{g,k}| \leq \epsilon_k \quad \forall i \in \{1, \dots, p\}, \\
|h(x^k)| &\leq \epsilon_k \quad \forall i \in \{1, \dots, m\}, \\
G_i(x^k) + \bar{t}_k &\geq -\epsilon_k, \nu_i^{G,k} \geq 0, \left| \nu_i^{G,k}(G_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad \forall i \in \{1, \dots, q\}, \\
H_i(x^k) + \bar{t}_k &\geq -\epsilon_k, \nu_i^{H,k} \geq 0, \left| \nu_i^{H,k}(H_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad \forall i \in \{1, \dots, q\}, \\
\Phi_i(G(x^k), H(x^k); t_k) &\leq \epsilon_k, \nu_i^{\Phi,k} \geq 0, \left| \nu_i^{\Phi,k} \Phi_i(G(x^k), H(x^k); t_k) \right| \leq \epsilon_k \quad \forall i \in \{1, \dots, q\}.
\end{aligned}$$

Unfortunately, it has been shown in [23] (Theorem 9 and 12) or [10] (Theorem 4.3) for the KDB, L-shape and butterfly relaxations that under this definition, sequences of epsilon-stationary points only converge to weak-stationary point without additional hypothesis. Our goal of computing an M-stationary point with a realistic method is far from obvious. Indeed, epsilon-stationary points have two main drawbacks considering our goal. The difficulties may come from the approximation of the complementarity condition and the approximate feasibility as shown in Example 5.1 or from the approximation of the feasibility of the relaxed constraint as illustrated in Example 5.2. In those examples, we consider the scheme (2) in order to simplify the presentation, but these observations can be easily generalized to the others methods.

Kanzow and Schwartz provide the following example exhibiting convergence to a W-stationary point.

Example 5.1. Consider the problem

$$\begin{aligned}
\min_{x \in \mathbb{R}^2} \quad & x_2 - x_1 \\
\text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0.
\end{aligned}$$

If we perturb the relation $\nu^\Phi \Phi(x_1, x_2; t) \leq \epsilon$ (leaving the other conditions $\nu^\Phi \geq 0, \Phi(x_1, x_2; t) \leq 0$), ν^Φ may be positive when the constraint $\Phi(x_1, x_2; t)$ is not active. For the case KDB $\Phi(x_1, x_2; t) = (x_1 - t)(x_2 - t) = -\epsilon^2$ with $\epsilon = t^2$, the point $x(t, \epsilon) = (t - \epsilon, t + \epsilon)^T \geq (0, 0)^T$ is epsilon-stationary for small enough ϵ : $\Phi(x_1, x_2; t) \leq \epsilon$ and the choice $\nu^\Phi = \frac{1}{\epsilon}$ makes the Lagrangian $(-1, 1)^T + \nu^\Phi \Phi(x_1, x_2; t)$ vanish. $x(t)$ converges to the origin when $t, \epsilon \rightarrow 0$ but the origin is only weakly stationary.

Now, if the complementarity constraint is relaxed, but the complementarity condition is guaranteed convergence may occur to C-stationary points as shown in the following example.

Example 5.2. Consider the problem

$$\begin{aligned}
\min_{x \in \mathbb{R}^2} \quad & \frac{1}{2}((x_1 - 1)^2 + (x_2 - 1)^2) \\
\text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0.
\end{aligned}$$

We specialize the relations in Definition 5.1 as the following, for t and ϵ close to 0.

$$\begin{aligned}
& \left\| \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} - \nu^G \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \nu^H \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \nu^\Phi \begin{pmatrix} x_2 - t \\ x_1 - t \end{pmatrix} \right\|_\infty \leq \epsilon, \\
& 0 \leq \nu^G, \quad (x_1 + t) \geq 0, \quad \nu^G(x_1 + t) \leq \epsilon, \\
& 0 \leq \nu^H, \quad (x_2 + t) \geq 0, \quad \nu^H(x_2 + t) \leq \epsilon, \\
& 0 \leq \nu^\Phi, \quad (x_1 - t)(x_2 - t) \leq \epsilon, \quad \nu^\Phi [(x_1 - t)(x_2 - t) - \epsilon] \geq 0.
\end{aligned}$$

The points $(t + \sqrt{\epsilon}, t + \sqrt{\epsilon})^T$ together with $\nu^G = \nu^H = 0$ and $\nu^\Phi = \frac{1-t-\sqrt{\epsilon}}{\sqrt{\epsilon}} \nearrow +\infty$ satisfy the above relations. The limit point when $t, \epsilon \rightarrow 0$ is the origin, which is a C-stationary point with $\nu^G = \nu^H = -1$.

On this example, the relaxed regularized complementarity constraint is active for any small enough $t, \epsilon > 0$; moreover, the relaxed regularized stationary point is a local maximum for $t + 2\sqrt{\epsilon} < 1$. The origin is a local maximum for the original (MPCC).

Another example might help understanding the phenomenon.

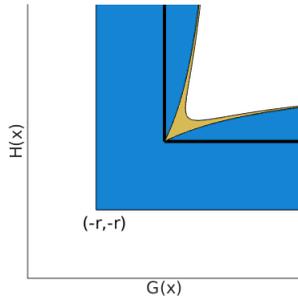


Figure 6: Butterfly relaxation with a constraint $\Phi^B(x; t) \leq \epsilon$.

Example 5.3. Consider the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & -\frac{1}{2}((x_1 - 1)^2 + (x_2 - 1)^2) \\ \text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0. \end{aligned}$$

We again specialize the relations in Definition 5.1 as the following, for t and ϵ close to 0.

$$\begin{aligned} & \left\| \begin{pmatrix} 1 - x_1 \\ 1 - x_2 \end{pmatrix} - \nu^G \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \nu^H \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \nu^\Phi \begin{pmatrix} x_2 - t \\ x_1 - t \end{pmatrix} \right\|_\infty \leq \epsilon, \\ & 0 \leq \nu^G, \quad (x_1 + t) \geq 0, \quad \nu^G(x_1 + t) \leq \epsilon, \\ & 0 \leq \nu^H, \quad (x_2 + t) \geq 0, \quad \nu^H(x_2 + t) \leq \epsilon, \\ & 0 \leq \nu^\Phi, \quad (x_1 - t)(x_2 - t) \leq \epsilon, \quad \nu^\Phi[(x_1 - t)(x_2 - t) - \epsilon] \geq 0. \end{aligned}$$

This time, the points $(t + \sqrt{\epsilon}, t + \sqrt{\epsilon})^T$ are no more epsilon-stationary but the points $x = (1, -t)^T$, $\nu^H = 1 + t$ and $x = (-t, 1)^T$, $\nu^G = 1 + t$ are. Their limits are $(1, 0)^T$ and $(0, 1)^T$ which are KKT points for the original MPCC with $\nu^H = 1, \nu^G = 0$ or $\nu^H = 0, \nu^G = 1$. The point $(-t, -t)^T$ with $\nu^H = 1 + t, \nu^G = 1 + t$ is also stationary, and of course converges to the origin, a local minimizer of the original MPCC.

In this example, the limit points are not minimizers for the original MPCC, but satisfy the first order KKT conditions for a minimizer. The second order conditions fails for those limit points. The two examples show limiting solutions of regularized subproblems which are not local minimizers of the original MPCC. The first one fails to satisfy a first order condition while the second one satisfies such a first order condition but not the second order one (it is a maximum on the active set).

The Figure 6 gives an intuition that explain the weak convergence in Example 5.2 by showing the ϵ -feasible set of the butterfly relaxed complementarity constraint. It can be noticed that this feasible set is very similar to the relaxation from Scheel and Scholtes, [28]. Therefore, it is no surprise that we can not expect more than convergence to a C-stationary point in these conditions.

6 Convergence of Epsilon-Stationary Sequences

We now address the convergence of sequences of epsilon-stationary points. This motivates the definition of a new kind of epsilon-stationary point called strong epsilon-stationary point, which is more stringent regarding the complementarity constraint.

Definition 6.1. x^k is a strong epsilon-stationary point for $(R_t(x))$ with $\epsilon_k \geq 0$ if there exists $\nu^k \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^{3q}$ such that

$$\|\nabla \mathcal{L}_{R_t}^1(x^k, \nu^k; t_k)\|_\infty \leq \epsilon_k$$

and

$$\begin{aligned}
g_i(x^k) &\leq \epsilon_k, \nu_i^{g,k} \geq 0, |g_i(x^k)\nu_i^{g,k}| \leq \epsilon_k \quad \forall i \in \{1, \dots, p\}, \\
|h(x^k)| &\leq \bar{t}_k + O(\epsilon_k) \quad \forall i \in \{1, \dots, m\}, \\
G_i(x^k) + \bar{t}_k &\geq -\epsilon_k, \nu_i^{G,k} \geq 0, \left| \nu_i^{G,k}(G_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad \forall i \in \{1, \dots, q\}, \\
H_i(x^k) + \bar{t}_k &\geq -\epsilon_k, \nu_i^{H,k} \geq 0, \left| \nu_i^{H,k}(H_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad \forall i \in \{1, \dots, q\}, \\
\Phi_i(G(x^k), H(x^k); t_k) &\leq 0, \nu_i^{\Phi,k} \geq 0, \left| \nu_i^{\Phi,k} \Phi_i(G(x^k), H(x^k); t_k) \right| = 0 \quad \forall i \in \{1, \dots, q\}.
\end{aligned}$$

In this case following a similar proof to the one of [23] and [10], we get an improved result that keep the nice properties of the relaxations without strong assumption on the sequence of $\{\epsilon_k\}$.

The following lemma shows that a sequence of strong epsilon-stationary points converges to a weak-stationary point. This is not a new result since the same has been proved in [23] and [10] for a sequence of epsilon-stationary points. However, the new definition allows to go further by showing convergence to an M-stationary point.

Lemma 6.1. *Given $\{t_k\}$ a sequence of parameter and $\{\epsilon_k\}$ a sequence of non-negative parameter such that both sequences decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \nu^k\}$ be a sequence of strong epsilon-stationary points of $(R_t(x))$ according to definition 6.1 for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$. Let $\{\eta^{G,k}\}, \{\eta^{H,k}\}$ be two sequences such that*

$$\begin{aligned}
\eta^{G,k} &:= \nu^{G,k} - \nu^{\Phi,k} \alpha^H(x^k; t_k), \\
\eta^{H,k} &:= \nu^{H,k} - \nu^{\Phi,k} \alpha^G(x^k; t_k).
\end{aligned} \tag{6}$$

Assume that the sequence of multipliers $\{\nu^{h,k}, \nu^{g,k}, \eta^{G,k}, \eta^{H,k}\}$ is bounded. Then, x^* is an M-stationary point of (MPCC).

Proof. The proof is divided in two parts. We first show that x^* is a weak-stationary point and then we prove that is an M-stationary point.

Let us prove the first part of the lemma. By definition $\{x^k, \nu^k\}$ is a sequence of strong epsilon-stationary points of $(R_t(x))$. We make the condition on the Lagrangian

$$\|\nabla \mathcal{L}_{R_t}^1(x^k, \nu^k; t_k)\|_\infty \leq \epsilon_k,$$

more explicit. By construction of $\Phi(G(x), H(x); t)$ this condition becomes

$$\|\nabla f(x^k) + \nabla g(x^k)^T \nu^{g,k} + \nabla h(x^k)^T \nu^{h,k} - \nabla G(x^k)^T \eta^{G,k} - \nabla H(x^k)^T \eta^{H,k}\|_\infty \leq \epsilon_k. \tag{7}$$

Besides, the sequence of multipliers $\{\nu^{h,k}, \nu^{g,k}, \eta^{G,k}, \eta^{H,k}\}$ is assumed bounded. Therefore, it follows that the sequence converges to some limit point

$$\{\nu^{h,k}, \nu^{g,k}, \eta^{G,k}, \eta^{H,k}\} \rightarrow (\nu^h, \nu^g, \eta^G, \eta^H).$$

It is to be noted that for k sufficiently large it holds

$$\begin{aligned}
\text{supp}(\nu^g) &\subset \text{supp}(\nu^{g,k}), \\
\text{supp}(\eta^G) &\subset \text{supp}(\eta^{G,k}), \\
\text{supp}(\eta^H) &\subset \text{supp}(\eta^{H,k}).
\end{aligned}$$

We prove that $(x^*, \nu^h, \nu^g, \eta^G, \eta^H)$ is a weak-stationary point. Obviously, since $\epsilon_k \downarrow 0$ it follows that $x^* \in \mathcal{Z}$, $\nabla_x \mathcal{L}_{MPCC}^1(x^*, \nu^h, \nu^g, \eta^G, \eta^H) = 0$ by (7) and that $\nu_i^g = 0$ for $i \notin \mathcal{I}_g(x^*)$. It remains to show that for indices

$i \in \mathcal{I}^{+0}(x^*)$, $\eta_i^G = 0$. The opposite case for indices $i \in \mathcal{I}^{+0}(x^*)$ would follow in a completely similar way. So, let i be in $\mathcal{I}^{+0}(x^*)$.

By definition of strong ϵ_k -stationarity it holds for all k that

$$|\nu_i^{G,k}(G_i(x^k) + \bar{t}_k)| \leq \epsilon_k.$$

Therefore, $\nu_i^{G,k} \rightarrow_{k \rightarrow \infty} 0$ since $\epsilon_k \downarrow 0$ and $G_i(x^k) \rightarrow G_i(x^*) > 0$.

Without loss of generality we may assume that for k sufficiently large $\nu_i^{\Phi,k} \neq 0$ otherwise $\eta_i^G = 0$ and the proof is complete. By strong ϵ -stationarity $\nu_i^{\Phi,k} \neq 0$ implies that $F_{H,i}(x^k; t_k) = 0$ by (H4). (H2) yields $\alpha_i^H(x^k; t_k) \rightarrow H_i(x^*)$ and so $\eta_i^G = 0$ unless $\nu^{\Phi,k}$ diverges as k grows. We now prove that the latter case leads to a contradiction.

Assume that $\nu^{\Phi,k} \rightarrow \infty$, boundedness hypothesis on η_i^G gives that there exists a finite non-vanishing constant C such that

$$\nu_i^{\Phi,k} \alpha_i^H(x^k; t_k) \rightarrow C.$$

Moreover, since η_i^H is finite and $\nu_i^{\Phi,k} \alpha_i^G(x^k; t_k) \rightarrow \infty$ as $G_i(x^k) > 0$ then necessarily $\nu_i^{H,k} \rightarrow \infty$. Furthermore, noticing that $F_H(x^k; t_k) = 0$ gives $H_i(x^k) \geq 0$, leads to a contradiction with $\nu_i^{H,k} \rightarrow \infty$ since by ϵ -stationarity we get

$$\left| \nu_i^{H,k}(H_i(x^k) + \bar{t}_k) \right| = \left| \nu_i^{H,k} H_i(x^k) \right| + \left| \nu_i^{H,k} \bar{t}_k \right| \leq \epsilon_k,$$

and $\epsilon_k = o(\bar{t}_k)$.

We can conclude that for $i \in \mathcal{I}^{+0}(x^*)$, $\eta_i^G = 0$ and therefore x^* is a weak-stationary point.

Now, let us prove that x^* is even stronger that weak-stationary point since it is an M-stationary point. We now consider indices $i \in \mathcal{I}^{00}(x^*)$. Our aim here is to prove that either $\eta_i^G > 0$, $\eta_i^H > 0$ or $\eta_i^G \eta_i^H = 0$. It is clear that if $\nu_i^{\Phi,k} = 0$, then η_i^G and η_i^H are non-negative values and the result holds true. So, without loss of generality we may assume that $\nu_i^{\Phi,k} \geq 0$ and then $\Phi_i(G(x^k), H(x^k); t_k) = 0$ by Definition 6.1.

By construction of $\Phi_i(G(x^k), H(x^k); t_k)$ given in hypothesis (H4) it follows that $\Phi_i(G(x^k), H(x^k); t_k) = 0 \iff F_{G,i}(x^k; t_k) = 0$ or $F_{H,i}(x^k; t_k) = 0$ where we remind that

$$\begin{aligned} F_{G,i}(x^k; t_k) &= G_i(x^k) - \psi(H_i(x^k); t_k), \\ F_{H,i}(x^k; t_k) &= H_i(x^k) - \psi(G_i(x^k); t_k), \end{aligned}$$

Without loss of generality we assume that $F_{G,i}(x^k; t_k) = 0$ since the other case is completely similar. Furthermore by construction of $\Phi_i(G(x^k), H(x^k); t_k)$ it holds that $G_i(x^k)$ is non-negative in this case.

Considering one of the complementarity condition of the strong ϵ -stationarity gives

$$\epsilon_k \geq |\nu_i^{G,k}(G_i(x^k) + \bar{t}_k)| = |\nu_i^{G,k} G_i(x^k)| + |\nu_i^{G,k} \bar{t}_k|,$$

since $G_i(x^k)$ is non-negative and it follows that

$$|\nu_i^{G,k} \bar{t}_k| \leq \epsilon_k.$$

Necessarily $\nu_i^{G,k} \rightarrow_{k \rightarrow \infty} 0$ as we assume in our statement that $\epsilon_k = o(\bar{t}_k)$.

Now at x^k we can use Lemma 3.1 that for $F_{G,i}(x^k; t_k) = 0$ gives

$$\begin{aligned} \alpha_i^G(x^k; t_k) &= - \frac{\partial \psi(x; t_k)}{\partial x} \Big|_{x=H_i(x^k)} F_{H,i}(x^k; t_k), \\ \alpha_i^H(x^k; t_k) &= F_{H,i}(x^k; t_k). \end{aligned}$$

Obviously, if $F_{H,i}(x^k; t_k) = 0$ we are done and so assume that $F_{H,i}(x^k; t_k) \neq 0$. By hypothesis (H6), it holds that $\frac{\partial \psi(x; t_k)}{\partial x} \Big|_{x=H_i(x^k)} \rightarrow_{k \rightarrow \infty} 0$. Therefore, $\alpha_i^G(x^k; t_k) \nu_i^{\Phi,k}$ going to a non-zero limit would imply that

$\alpha_i^H(x^k; t_k) \nu_i^{\Phi, k}$ goes to infinity. However, this is a contradiction with η_i^G being finite. We can conclude that necessarily $\alpha_i^G(x^k; t_k) \nu_i^{\Phi, k}$ converges to zero.

Finally, we examine two cases regarding the sign of $F_{H_i}(x^k; t_k)$. For $F_{H_i}(x^k; t_k) \leq 0$, we get η_i^G, η_i^H non-negative, which satisfy the desired condition. For $F_{H_i}(x^k; t_k) \geq 0$ we get $\nu_i^{H, k} \rightarrow_{k \rightarrow \infty} 0$ using the same argument than for $\nu_i^{G, k}$. Thus, it follows that $\eta_i^H = 0$.

This concludes the proof that x^* is an M-stationary point, since additionally to the proof of weak-stationarity of x^* we proved for every $i \in \mathcal{I}^{00}(x^*)$ that either $\eta_i^H > 0, \eta_i^G > 0$ or $\eta_i^H \eta_i^G = 0$. \square

The following theorem is a direct consequence of both previous lemmas and is our main statement.

Theorem 6.1. *Given $\{t_k\}$ a sequence of parameter and $\{\epsilon_k\}$ a sequence of non-negative parameter such that both sequences decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \nu^k\}$ be a sequence of epsilon-stationary points of $(R_t(x))$ according to definition 6.1 for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-CRSC holds at x^* . Then, x^* is an M-stationary point of (MPCC).*

Proof. The proof is direct by Lemma 6.1 and Corollary 2.3 of [10] that ensures boundedness of the sequence (6) under MPCC-CRSC. \square

Theorem (6.1) attains the ultimate goal, however it is not a trivial task to compute such a sequence of epsilon-stationary points. This is discussed later. Another important question is the existence of strong epsilon-stationary points in the neighbourhood of an M-stationary point. This problem is tackled in the following sections.

7 On Lagrange Multipliers of the Regularization

The following example develops on Example 2.1 due to Kanzow and Schwartz exhibits a situation where the regularized subproblems have no KKT point.

Example 7.1. *The KDB regularized problem is*

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & x_1 + x_2 - x_3 \\ \text{s.t.} \quad & -4x_1 + x_3 \leq 0, \\ & -4x_2 + x_3 \leq 0, \\ & x_1 \geq -t, \\ & x_2 \geq -t, \\ & (x_1 - t)(x_2 - t) \leq 0. \end{aligned} \tag{8}$$

The point $(t, t, 4t)^t$ is feasible so that the minimum value of this program is $\leq -2t$. Moreover, whenever $x_1 > t$, we must have $x_2 \leq t$ to satisfy $(x_1 - t)(x_2 - t) \leq 0$. This allows to conclude that $(t, t, 4t)^t$ is the global minimum of the regularized problem. $\nu^G = \nu^H = 0$ and the gradient of the Lagrangian equal to zero yields

$$0 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \nu_1^g \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + \nu_2^g \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} + \nu^\Phi \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{9}$$

which cannot be satisfied.

This last example seems to contradict Theorem 4.6 in [22], but MPCC-LICQ is not satisfied by four constraints in \mathbb{R}^3 .

It has been pointed out earlier that a practical algorithm may not be able to compute stationary point of the regularized subproblem, but only some approximate epsilon-stationary point. An intuitive idea would be that weaker constraint qualification may guarantee existence of such points. However, the following one dimensional example shows that things are not that simple taking for instance the approximate method KDB, (2).

Example 7.2. Consider the problem

$$\min_{x \in \mathbb{R}} -x, \text{ s.t } 0 \leq G(x) := x \perp H(x) := 0 \geq 0.$$

G and H are linear functions, so MPCC-CRCQ holds at each feasible point. Clearly, $x = 0$ is an M-stationary point with $\lambda^G = -1$ and $\lambda^H = 0$. Indeed, the gradient of the Lagrangian is given by

$$0 = -1 - \lambda^G.$$

We now verify that there is no epsilon-stationary point of the approximation KDB for this example :

$$\min_{x \in \mathbb{R}} -x, \text{ s.t } x \geq -t, -t(x - t) \leq 0.$$

Now, considering that the gradient of the Lagrangian function for this problem must be lower or equal as ϵ gives

$$\| -1 - \nu^G - t\nu^\Phi \| \leq \epsilon.$$

Noticing that $\nu^G, \nu^H, \nu^\Phi \geq 0$ leads to

$$1 + \nu^G + t\nu^\Phi \leq \epsilon,$$

which leads to a contradiction for $\epsilon < 1$. So, there is no sequence of approximate stationary points that goes to the origin.

The previous example illustrates the fact that even so strong constraint qualification holds for the problem existence of epsilon-stationary point are not ensured at an M-stationary point. Even so, this problem seems intractable by reformulating the (MPCC) with slack variables things could be slightly different.

Example 7.3. (Example 7.2 continued) We now verify that there is strong epsilon-stationary point of the approximation KDB written with slack variables for this example :

$$\min_{x \in \mathbb{R}} -x, \text{ s.t } s_G = x, s_H = 0, s_G \geq -t, s_H \geq -t, (s_G - t)(s_H - t) \leq 0.$$

Given $\delta > 0$, consider the point $x = 0, (s_G, s_H) = (t, t + \delta)$ and Lagrange multiplier $(\nu^{s_G}, \nu^{s_H}, \nu^G, \nu^H, \nu^\Phi) = (-1, 0, 0, 0, \frac{1}{\delta})$.

- Condition on the gradient of the Lagrangian

$$\begin{aligned} |\nabla_x \mathcal{L}^1(x, s, \nu)| &= |-1 - \nu^{s_G}| = 0, \\ |\nabla_{s_G} \mathcal{L}^1(x, s, \nu)| &= |\nu^{s_G} - \nu^G + \nu^\Phi(s_H - t)| = 0, \\ |\nabla_{s_H} \mathcal{L}^1(x, s, \nu)| &= |\nu^{s_H} - \nu^H + \nu^\Phi(s_G - t)| = 0. \end{aligned}$$

- Condition on the feasibility

$$\begin{aligned} |x - s_G| &= t \leq \epsilon, \\ |0 - s_H| &= t + \delta \leq \epsilon, \\ s_G + t &= 2t \geq -\epsilon, \quad s_H + t = 2t + \delta \geq -\epsilon, \\ (s_G - t)(s_H - t) &= 0. \end{aligned}$$

- Condition on the complementarity

$$\begin{aligned} |(s_G + t)\nu^G| &= 0, \quad |(s_H + t)\nu^H| = 0, \\ |(s_G - t)(s_H - t)\nu^\Phi| &= 0. \end{aligned}$$

This completes the proof that there is a strong epsilon-stationary point for the formulation with slack variables.

This example motivates the use of slack variables to define the (MPCC) and in this case study the existence of strong epsilon-stationary point in a neighbourhood of an M-stationary point.

8 The MPCC with Slack Variables

Consider the following parametric non-linear program $R_t(x, s)$ parametrized by t :

$$\begin{aligned} \min_{(x,s) \in \mathbb{R}^n \times \mathbb{R}^{2q}} & f(x) \\ \text{s.t.} & g(x) \leq 0, \quad h(x) = 0, \\ & s_G = G(x), \quad s_H = H(x), \\ & s_G \geq -\bar{t}\mathbf{1}, \quad s_H \geq -\bar{t}\mathbf{1}, \quad \Phi(s_G, s_H; t) \leq 0, \end{aligned} \quad (R_t^s(x, s))$$

with $\lim_{\|t\| \rightarrow 0} \bar{t} = 0^+$ and the relaxation map $\Phi(s_G, s_H; t) : \mathbb{R}^q \times \mathbb{R}^q \rightarrow \mathbb{R}^q$ is defined by replacing $G(x)$ and $H(x)$ by s_G and s_H in the map $\Phi(G(x), H(x); t)$.

The generalized Lagrangian function of $(R_t^s(x, s))$ is defined as

$$\begin{aligned} \mathcal{L}_s^r(x, s, \nu; t) := & r f(x) + g(x)^T \nu^g + h(x)^T \nu^h - (G(x) - s_G)^T \nu^{s_G} - (H(x) - s_H)^T \nu^{s_H} \\ & - s_G^T \nu^G - s_H^T \nu^H + \Phi(s_G, s_H; t)^T \nu^\Phi. \end{aligned}$$

Let \mathcal{F}_s be the feasible set of $(R_t^s(x, s))$.

The following result is a direct corollary of Theorem 6.1 stating that the reformulation with slack variables does not alter the convergence result.

Corollary 8.1. *Given $\{t_k\}$ a sequence of parameter and $\{\epsilon_k\}$ a sequence of non-negative parameter such that both sequences decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \nu^k\}$ be a sequence of strong epsilon-stationary points of $(R_{t_k}^s(x, s))$ for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-CRSC holds at x^* . Then, x^* is an M-stationary point of (MPCC).*

Proof. Let $\tilde{h}(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$ be such that $\tilde{h}(x) := (h(x), s_G - G(x), s_H - H(x))$ and $\tilde{x} := (x, s_G, s_H)$. It is clear that the non-linear program $(R_t^s(x, s))$ fall under the formulation $(R_t(x))$. Therefore, we can apply 6.1 to conclude this proof. \square

The following lemma giving an explicit form of the gradient of the Lagrangian function of $(R_t^s(x, s))$ can be deduced through direct computations.

Lemma 8.1. *The gradient of $\mathcal{L}_s^r(x, s, \nu; t)$ is given by*

$$\nabla_x \mathcal{L}_s^r(x, s, \nu; t) = r \nabla f(x) + \nabla g(x)^T \nu^g + \nabla h(x)^T \nu^h - \nabla G(x)^T \nu^{s_G} - \nabla H(x)^T \nu^{s_H}, \quad (10)$$

$$\begin{aligned} \nabla_{s_G} \mathcal{L}_s^r(x, s, \nu; t) &= \nu^{s_G} - \nu^G + \nabla_{s_G} \Phi(s_G, s_H; t)^T \nu^\Phi, \\ \nabla_{s_H} \mathcal{L}_s^r(x, s, \nu; t) &= \nu^{s_H} - \nu^H + \nabla_{s_H} \Phi(s_G, s_H; t)^T \nu^\Phi. \end{aligned} \quad (11)$$

There is two direct consequences of this result. First, it is easy to see from this lemma that computing a stationary point of $\mathcal{L}_s^r(x, s, \nu; t)$ is equivalent to computing a stationary point of $\mathcal{L}^r(x, \nu; t)$. Secondly, a stationary point of $\mathcal{L}_s^r(x, s, \nu; t)$ with $r = 1$ satisfies one of the condition of weak-stationary point of (MPCC) that is $\nabla \mathcal{L}_{MPCC}^1(x, \nu) = 0$.

In the next section, we now consider the existence of strong epsilon-stationary point for the relaxation with slack variables $(R_t^s(x, s))$.

9 Existence of Strong Epsilon-Stationary Points for the Regularization with Slack Variables

Before stating our main result we give a serie of additional hypothesis on the relaxation and the function ψ . It is essential to note once again, that all these hypothesis are not restrictive, since they are satisfied by the existing methods in the literature.

9.1 Assumptions

The assumptions made in this section are divided in two parts. The first part concerns assumptions on the domain of the relaxation. The second part is assumptions on the relaxation function ψ that has been used to define the relaxation map on the boundary of the feasible set in Section 3.

9.1.1 Assumptions on the Relaxations

We denote $\mathbb{B}_{c\epsilon}((-\bar{t}\mathbf{1}, -\bar{t}\mathbf{1})^T)$ the ball of radius $c\epsilon$ and centered in $(-\bar{t}\mathbf{1}, -\bar{t}\mathbf{1})^T$. We assume that the schemes considered in the sequel satisfy belong to one of two following cases for positive constants c, ϵ and \bar{t} :

Case 1

$$\mathbb{B}_{c\epsilon}((-\bar{t}\mathbf{1}, -\bar{t}\mathbf{1})^T) \cap \{(s_G, s_H)^T \mid s_G \geq -\bar{t}\mathbf{1}, s_H \geq -\bar{t}\mathbf{1}\} \subset \mathcal{F}_s; \quad (\text{F1})$$

Case 2

$$(\mathbb{B}_{c\epsilon}((-\bar{t}\mathbf{1}, \psi(-\bar{t}\mathbf{1}; t))^T) \cup \mathbb{B}_{c\epsilon}((\psi(-\bar{t}\mathbf{1}; t), -\bar{t})^T)) \cap \{(s_G, s_H)^T \mid s_G \geq -\bar{t}\mathbf{1}, s_H \geq -\bar{t}\mathbf{1}, \Phi(s_G, s_H; t) = 0\} \subset \mathcal{F}_s. \quad (\text{F2})$$

The first case includes the butterfly relaxation and the KS relaxation, while the second case includes the approximation KDB.

9.1.2 Assumptions on the Relaxation Function

For all $t \in \mathbb{R}_{++}^l$, we make the following supplementary assumptions on the function ψ for all $x \in \mathbb{R}^q$. We remind here that the functions ψ are separable with respect to x .

- $$\frac{\partial \psi(x; t)}{\partial t} > 0; \quad (\text{A1})$$

- $$\frac{\partial \psi(x; t)}{\partial x} \geq 0; \quad (\text{A2})$$

- $$\psi(\psi(\|t\|_\infty; t); t) \leq \|t\|_\infty; \quad (\text{A3})$$

- $$\psi(-\|t\|_\infty; t) \leq \|t\|_\infty. \quad (\text{A4})$$

Hypothesis (A1) in particular implies some monotonicity on the feasible set of the relaxed problems. Assumption (A4) is used for the second kind of relaxations only. It is to be noted here that the assumptions (A1),(A2),(A3) and (A4) are not the weakest for obtaining the following results. However, those assumptions are satisfied by all the relaxations defined in the literature.

Lemma 9.1. *Assume that (A1), (A2) and (A3) hold true. Then, giving constants $c > 0$, $\bar{t} > 0$ and $\epsilon > 0$ the following holds true for all $\|t\|_\infty \in (0, \bar{t} + c\epsilon)$*

$$\bar{t} + c\epsilon - \psi(\psi(\bar{t} + c\epsilon; t); t) > 0.$$

Proof. Using (A1), (A2) and that $\|t\|_\infty \in (0, \bar{t} + c\epsilon)$ yields

$$\bar{t} + c\epsilon - \psi(\psi(\bar{t} + c\epsilon; t); t) > \bar{t} + c\epsilon - \psi(\psi(\bar{t} + c\epsilon; e(\bar{t} + c\epsilon)); e(\bar{t} + c\epsilon)) \geq 0.$$

The conclusion comes from assumption (A3). □

Lemma 9.2. *Assume that (A1) and (A3) holds true. Then, giving constants $c > 0$, $\bar{t} > 0$ and $\epsilon > 0$ the following holds true for all $\|t\|_\infty \in (0, \bar{t} + c\epsilon]$*

$$\psi(\bar{t} + c\epsilon; t) \leq \bar{t} + c\epsilon.$$

Proof. Using assumption (A1) and then (A3) gives

$$\psi(\bar{t} + c\epsilon; t) < \psi(\bar{t} + c\epsilon; e(\bar{t} + c\epsilon)) \leq \bar{t} + c\epsilon,$$

which concludes the proof. \square

Lemma 9.3. *Given positive constants \bar{t}, c, ϵ, K . There exists a $t^* > 0$ such that for all $t \in (0, t^*]$ it holds that*

$$\left. \frac{\partial \psi(x; t)}{\partial x} \right|_{x=\bar{t}+c\epsilon} \leq K\epsilon,$$

and

$$0 \leq \bar{t} - \left. \frac{\partial \psi(x; t)}{\partial x} \right|_{x=c\epsilon} \leq K\epsilon.$$

Proof. The proof is clear from Assumption (H6) on the relaxation. \square

Lemma 9.4. *Assume that (A1) and (A4) holds true. Then, giving constants $c > 0$, $\bar{t} > 0$ and $\epsilon > 0$ with $\bar{t} > c\epsilon$ the following holds true for all $\|t\|_\infty \in (0, \bar{t} - c\epsilon]$*

$$\psi(\bar{t} - c\epsilon; t) \leq \bar{t} + c\epsilon.$$

Proof. Using assumption (A1) and then (A4) gives

$$\psi(\bar{t} - c\epsilon; t) < \psi(\bar{t} - c\epsilon; e(\bar{t} - c\epsilon)) \leq \bar{t} - c\epsilon \leq \bar{t} + c\epsilon,$$

which concludes the proof. \square

9.2 Main Theorem on Existence of Lagrange Multiplier

All of the supplementary assumptions made above are now used to derive the following result.

Theorem 9.1. *Let $x^* \in \mathcal{Z}$ be an M-stationary point and $\epsilon > 0$ be arbitrarily small. Furthermore, assume that the hypothesis (A1),(A2),(A3),(A4) on ψ and the hypothesis (F1) or (F2) on the relaxation introduced above hold true. Then, there exists positive constants c , t^* and \bar{t}^* with $\bar{t}^* > c\epsilon$ and a neighbourhood $U(x^*)$ of $(x^*, 0)^T$ such that for all $t \in (0, t^*)$ and $\bar{t} \in (0, \bar{t}^*)$ there exists $(x, s)^T \in U(x^*)$, which is strong epsilon-stationary point of the relaxation $(R_i^s(x, s))$.*

Regarding the value of t^* we need at least that $\|t\|_\infty \leq \bar{t} - c\epsilon$. The constant c is given in the proof and depends on the multipliers of the M-stationary point.

Proof. The proof is conducted in two steps. First, we construct a point based on the solution that is a candidate to be a strong epsilon-stationary point. Then, we verify that this candidate is actually a strong epsilon-stationary point.

x^* is assumed to be an M-stationary point. Therefore, there exists $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ such that

$$\begin{aligned} \nabla \mathcal{L}_{MPCC}^1(x^*, \lambda) &= 0, \\ \min(\lambda^g, -g_i(x^*)) &= 0, \quad h(x^*) = 0, \quad \lambda_{\mathcal{I}^0(x^*)}^G = 0, \quad \lambda_{\mathcal{I}^0(x^*)}^H = 0, \\ \text{either } \lambda_i^G > 0, \lambda_i^H > 0 &\text{ either } \lambda_i^G \lambda_i^H = 0 \text{ for } i \in \mathcal{I}^{00}(x^*). \end{aligned}$$

Let c be the positive constant bounding the value of the Lagrange multipliers so that

$$c := \max_{i \in \text{supp}(\lambda^G), j \in \text{supp}(\lambda^H)} \frac{1}{|\lambda_i^G|} + \frac{1}{|\lambda_j^H|}.$$

Construction of the point $(\hat{x}, \hat{s}, \hat{\nu})$ Let us construct a point $(\hat{x}, \hat{s}, \hat{\nu})$ that satisfy the strong epsilon-stationary conditions (6.1).

$$\hat{x} := x^*, \quad \hat{\nu}^g := \lambda^g, \quad \hat{\nu}^h := \lambda^h, \quad \hat{\nu}^{sG} := \lambda^G, \quad \hat{\nu}^{sH} := \lambda^H.$$

We now split into two cases **(A)** and **(B)** corresponding to the two different kind of relaxations. Denote the following set of indices

$$\begin{aligned} \mathcal{I}_{-0}^{00}(x^*, \lambda) &:= \{i \in \{1, \dots, q\} \mid \lambda_i^G < 0, \lambda_i^H = 0\}, \\ \mathcal{I}_{0-}^{00}(x^*, \lambda) &:= \{i \in \{1, \dots, q\} \mid \lambda_i^G = 0, \lambda_i^H < 0\}, \\ \mathcal{I}_{\nu^G} &:= \text{supp}(\hat{\nu}^{sG}) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{0-}^{00}(x^*, \lambda)), \\ \mathcal{I}_{\nu^H} &:= \text{supp}(\hat{\nu}^{sH}) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{0-}^{00}(x^*, \lambda)), \end{aligned}$$

A) Consider the Case 1, we choose $\hat{s}_G, \hat{s}_H, \hat{\nu}^G, \hat{\nu}^H$ and $\hat{\nu}^\Phi$ such that :

$$\begin{aligned} \hat{s}_G &:= \begin{cases} \psi(\bar{t} + c\epsilon; t), & i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \bar{t} + c\epsilon, & i \in \mathcal{I}_{0-}^{00}(x^*, \lambda), \\ \frac{\epsilon - \bar{t}\hat{\nu}_i^G}{\hat{\nu}_i^G}, & i \in \mathcal{I}_{\nu^G}, \\ \in \mathcal{F} & \text{otherwise,} \end{cases} \\ \hat{s}_H &:= \begin{cases} \bar{t} + c\epsilon, & i \in \mathcal{I}_{0-}^{00}(x^*, \lambda), \\ \psi(\bar{t} + c\epsilon; t), & i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \frac{\epsilon - \bar{t}\hat{\nu}_i^H}{\hat{\nu}_i^H}, & i \in \mathcal{I}_{\nu^H}, \\ \in \mathcal{F} & \text{otherwise,} \end{cases} \\ \hat{\nu}^G &:= \begin{cases} \hat{\nu}_i^{sG} & \text{for } i \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{0-}^{00}(x^*, \lambda)), \\ 0 & \text{otherwise,} \end{cases} \\ \hat{\nu}^H &:= \begin{cases} \hat{\nu}_i^{sH} & \text{for } i \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{0-}^{00}(x^*, \lambda)), \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and finally

$$\hat{\nu}^\Phi := \begin{cases} \frac{-\hat{\nu}_i^{sG}}{\alpha_i^H(s; t)}, & \text{for } i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \frac{-\hat{\nu}_i^{sH}}{\alpha_i^G(s; t)}, & \text{for } i \in \mathcal{I}_{0-}^{00}(x^*, \lambda), \\ 0, & \text{otherwise.} \end{cases}$$

$(\hat{x}, \hat{s}, \hat{\nu})$ **satisfy the stationarity** : Finally, we verify that in both cases we satisfy the strong epsilon-stationary conditions, that is

$$\|\nabla \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t)\|_\infty \leq \epsilon,$$

and

$$\begin{aligned} g_i(\hat{x}) &\leq \epsilon, \quad \hat{\nu}_i^g \geq 0, \quad |g_i(\hat{x})\hat{\nu}_i^g| \leq \epsilon \quad \forall i \in \{1, \dots, p\}, \\ |h_i(\hat{x})| &\leq \bar{t} + c\epsilon \quad \forall i \in \{1, \dots, m\}, \\ |G_i(\hat{x}) - \hat{s}_{G,i}| &\leq \bar{t} + c\epsilon \quad \forall i \in \{1, \dots, q\}, \\ |H_i(\hat{x}) - \hat{s}_{H,i}| &\leq \bar{t} + c\epsilon \quad \forall i \in \{1, \dots, q\}, \\ \hat{s}_{G,i} + \bar{t} &\geq -\epsilon, \quad \hat{\nu}_i^G \geq 0, \quad |\hat{\nu}_i^G(\hat{s}_{G,i} + \bar{t})| \leq \epsilon \quad \forall i \in \{1, \dots, q\}, \\ \hat{s}_{H,i} + \bar{t} &\geq -\epsilon, \quad \hat{\nu}_i^H \geq 0, \quad |\hat{\nu}_i^H(\hat{s}_{H,i} + \bar{t})| \leq \epsilon \quad \forall i \in \{1, \dots, q\}, \\ \Phi_i(\hat{s}_G, \hat{s}_H; t) &\leq 0, \quad \hat{\nu}_i^\Phi \geq 0, \quad |\hat{\nu}_i^\Phi \Phi_i(\hat{s}_G, \hat{s}_H; t)| \leq 0 \quad \forall i \in \{1, \dots, q\}. \end{aligned}$$

We split the rest of the proof of **(A)** in 6 parts:

- I. $\|\nabla_x \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t)\|_\infty \leq \epsilon$, $g_i(\hat{x}) \leq \epsilon$, $\hat{\nu}_i^g \geq 0$, $|g_i(\hat{x})\hat{\nu}_i^g| \leq \epsilon \forall i \in \{1, \dots, p\}$ and $|h_i(\hat{x})| \leq \bar{t} + c\epsilon \forall i \in \{1, \dots, m\}$;
- II. $\|\nabla_s \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t)\|_\infty \leq \epsilon$;
- III. $|G_i(\hat{x}) - \hat{s}_{G,i}| \leq \bar{t} + c\epsilon$, $|H_i(\hat{x}) - \hat{s}_{H,i}| \leq \bar{t} + c\epsilon \forall i \in \{1, \dots, q\}$;
- IV. $\hat{s}_{G,i} + \bar{t} \geq -\epsilon$, $|\hat{\nu}_i^G(\hat{s}_{G,i} + \bar{t})| \leq \epsilon$, $\hat{s}_{H,i} + \bar{t} \geq -\epsilon$, $|\hat{\nu}_i^H(\hat{s}_{H,i} + \bar{t})| \leq \epsilon \forall i \in \{1, \dots, q\}$;
- V. $\Phi_i(\hat{s}_G, \hat{s}_H; t) \leq 0$, $|\hat{\nu}_i^\Phi \Phi_i(\hat{s}_G, \hat{s}_H; t)| \leq 0 \forall i \in \{1, \dots, q\}$;
- VI. $\hat{\nu}_i^G \geq 0$, $\hat{\nu}_i^H \geq 0$, $\hat{\nu}_i^\Phi \geq 0$.

Let us now run through these 6 conditions.

I. Since, $\hat{x} = x^*$ and $(\hat{\nu}^g, \hat{\nu}^h, \nu^{s_G}, \nu^{s_H}) = (\lambda^g, \lambda^h, \lambda^G, \lambda^H)$ it holds that

$$\|\nabla_x \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t)\|_\infty = 0,$$

and

$$\begin{aligned} g_i(\hat{x}) &\leq 0, \quad \hat{\nu}_i^g \geq 0, \quad |g_i(\hat{x})\hat{\nu}_i^g| \leq 0 \quad \forall i \in \{1, \dots, p\}, \\ |h_i(\hat{x})| &= 0 \quad \forall i \in \{1, \dots, m\}. \end{aligned}$$

II. The gradient of the Lagrangian with respect to s is given by

$$\begin{aligned} \nabla_{s_G} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}^{s_G} - \hat{\nu}^G + \hat{\nu}^\Phi \alpha^H(\hat{s}; t), \\ \nabla_{s_H} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}^{s_H} - \hat{\nu}^H + \hat{\nu}^\Phi \alpha^G(\hat{s}; t). \end{aligned}$$

In the case $\mathcal{I}_{-0}^{00}(x^*, \lambda)$ (the case $\mathcal{I}_{+0}^{00}(x^*, \lambda)$ is similar by symmetry) it is true that $\hat{\nu}_i^G = \hat{\nu}_i^H = 0$ and $\hat{\nu}_i^\Phi = \frac{-\hat{\nu}_i^{s_G}}{\alpha_i^H(\hat{s}; t)}$. Therefore

$$\begin{aligned} \nabla_{s_G, i} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= 0, \\ \nabla_{s_H, i} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \frac{-\hat{\nu}_i^{s_G} \alpha_i^G(\hat{s}; t)}{\alpha_i^H(\hat{s}; t)} = \hat{\nu}_i^{s_G} \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=\bar{t}+c\epsilon}, \end{aligned}$$

since for $i \in \mathcal{I}_{-0}^{00}(x^*, \lambda)$ by construction of \hat{s}_G, \hat{s}_H it holds that $F_{G_i} = 0$ and $\hat{s}_{H,i} = \bar{t} + c\epsilon$. The conclusion follows by Lemma 9.3, which gives

$$\frac{\partial \psi(x; t)}{\partial x} \Big|_{x=\bar{t}+c\epsilon} \leq \epsilon.$$

Now, in the cases $\mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{+0}^{00}(x^*, \lambda))$ and $\mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*) \setminus (\mathcal{I}_{-0}^{00}(x^*, \lambda) \cup \mathcal{I}_{+0}^{00}(x^*, \lambda))$ the construction of the multipliers gives directly that $\nabla_{(s_G, s_H)} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) = 0$.

This concludes the proof of **II**.

III. \hat{x} feasible for the MPCC yields to

$$\begin{aligned} |G_i(\hat{x}) - \hat{s}_{G,i}| &= |\hat{s}_{G,i}| \quad \text{and} \quad |H_i(\hat{x}) - \hat{s}_{H,i}| = 0, \quad \text{for } i \in \mathcal{I}^{0+}(x^*), \\ |G_i(\hat{x}) - \hat{s}_{G,i}| &= 0 \quad \text{and} \quad |H_i(\hat{x}) - \hat{s}_{H,i}| = |\hat{s}_{H,i}|, \quad \text{for } i \in \mathcal{I}^{+0}(x^*), \\ |G_i(\hat{x}) - \hat{s}_{G,i}| &= |\hat{s}_{G,i}| \quad \text{and} \quad |H_i(\hat{x}) - \hat{s}_{H,i}| = |\hat{s}_{H,i}|, \quad \text{for } i \in \mathcal{I}^{00}(x^*). \end{aligned}$$

By symmetry it is sufficient to consider the variables s_G . We analyse the cases where $i \in \mathcal{I}_{-0}^{00}(x^*, \lambda)$, $\mathcal{I}_{+0}^{00}(x^*, \lambda)$ and \mathcal{I}_{ν^G} .

- Let $i \in \mathcal{I}_{-0}^{00}(x^*, \lambda)$, then $\hat{s}_{G,i} = \bar{t} + c\epsilon$;
- Let $i \in \mathcal{I}_{+0}^{00}(x^*, \lambda)$, then $\hat{s}_{G,i} = \psi(\bar{t} + c\epsilon; t) \leq \bar{t} + c\epsilon$ by Lemma 9.2;

- Let $i \in \mathcal{I}_{\nu^G}$, then $\hat{s}_{G,i} = \left| \frac{\epsilon}{\hat{\nu}_i^G} - \bar{t} \right| \leq \left| \frac{\epsilon}{\bar{\nu}_i^G} \right| + \bar{t} \leq \bar{t} + c\epsilon$.

In every cases it holds that $|\hat{s}_{G,i}| \leq \bar{t} + c\epsilon$ and so this III is verified.

IV. By construction $\hat{s}_{G,i}$ and $\hat{s}_{H,i}$ are both non-negative as $\psi(\cdot; t)$ is assumed non-negative for indices i such that $\Phi_i(\hat{s}_G, \hat{s}_H; t) = 0$.

It remains to verify the condition in the case where $i \in \mathcal{I}_{\nu^G}$ and $i \in \mathcal{I}_{\nu^H}$. However, in both cases it holds that

$$\begin{aligned} \forall i \in \mathcal{I}_{\nu^G}, \quad \hat{s}_{G,i} + \bar{t} &= \frac{\epsilon - \bar{t}\hat{\nu}_i^G}{\hat{\nu}_i^G} + \bar{t} = \frac{\epsilon}{\hat{\nu}_i^G} > 0 \geq -\epsilon, \\ \forall i \in \mathcal{I}_{\nu^H}, \quad \hat{s}_{H,i} + \bar{t} &= \frac{\epsilon - \bar{t}\hat{\nu}_i^H}{\hat{\nu}_i^H} + \bar{t} = \frac{\epsilon}{\hat{\nu}_i^H} > 0 \geq -\epsilon. \end{aligned}$$

So, the feasibility in condition IV is satisfied. Now, regarding the complementarity condition it holds that

$$\begin{aligned} \forall i \in \mathcal{I}_{\nu^G}, \quad |(\hat{s}_{G,i} + \bar{t})\hat{\nu}_i^G| &= \left| \left(\frac{\epsilon - \bar{t}\hat{\nu}_i^G}{\hat{\nu}_i^G} + \bar{t} \right) \hat{\nu}_i^G \right| = \epsilon, \\ \forall i \in \mathcal{I}_{\nu^H}, \quad |(\hat{s}_{H,i} + \bar{t})\hat{\nu}_i^H| &= \left| \left(\frac{\epsilon - \bar{t}\hat{\nu}_i^H}{\hat{\nu}_i^H} + \bar{t} \right) \hat{\nu}_i^H \right| = \epsilon. \end{aligned}$$

This proves that the complementarity condition holds true for the relaxed positivity constraints and so condition IV is verified.

V. The feasibility $\Phi_i(\hat{s}_G, \hat{s}_H; t) \leq 0$ and the complementarity condition $|\hat{\nu}_i^\Phi \Phi_i(\hat{s}_G, \hat{s}_H; t)| \leq 0$ are satisfied by construction and by hypothesis on the relaxation.

VI. The multiplier $\hat{\nu}^\Phi$ is non-negative since for $i \in \mathcal{I}_{-0}^{00}(x^*, \lambda)$ it holds that

$$\alpha_i^H(\hat{s}; t) = F_{H_i}(\hat{s}; t) - \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=s_{G,i}} F_{G_i}(\hat{s}; t) = F_{H_i}(\hat{s}; t) > 0, \quad (12)$$

since $F_{G_i}(\hat{s}; t) = 0$ by construction of $\hat{s}_{G,i}$ and $F_H(\bar{t} + c\epsilon; t) > 0$ by Lemma 9.1. The case $i \in \mathcal{I}_{0-}^{00}(x^*, \lambda)$ follows by symmetry.

The others multipliers are obviously non-negative by construction. This concludes the case VI.

The verification of all 6 cases proves that the point constructed above is strong epsilon-stationary, which concludes the proof of the relaxations **(A)**.

B) Consider the Case 2. Let $\hat{s}_G, \hat{s}_H, \hat{\nu}^G, \hat{\nu}^H$ and $\hat{\nu}^\Phi$ be such that :

$$\hat{s}_G := \begin{cases} \psi(\bar{t} + c\epsilon; t), & i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \bar{t} + c\epsilon, & i \in \mathcal{I}_{0-}^{00}(x^*, \lambda), \\ \frac{\epsilon - \bar{t}\hat{\nu}_i^G}{\hat{\nu}_i^G}, & i \in \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}, \\ \psi\left(\frac{\epsilon - \bar{t}\hat{\nu}_i^H}{\hat{\nu}_i^H}; t\right), & i \in \mathcal{I}_{\nu^H} \setminus \mathcal{I}_{\nu^G}, \\ \in \mathcal{F} & \text{otherwise,} \end{cases} \quad (13)$$

$$\hat{s}_H := \begin{cases} \bar{t} + c\epsilon, & i \in \mathcal{I}_{0-}^{00}(x^*, \lambda), \\ \psi(\bar{t} + c\epsilon), & i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \psi\left(\frac{\epsilon - \bar{t}\hat{\nu}_i^G}{\hat{\nu}_i^G}; t\right), & i \in \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}, \\ \frac{\epsilon - \bar{t}\hat{\nu}_i^H}{\hat{\nu}_i^H}, & i \in I_2, \\ \in \mathcal{F} & \text{otherwise,} \end{cases} \quad (14)$$

$$\hat{\nu}^G := \begin{cases} \hat{\nu}_i^{sG} & \text{for } i \in \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}, \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

$$\hat{\nu}^H := \begin{cases} \hat{\nu}_i^{sH} & \text{for } i \in \mathcal{I}_{\nu^H} \setminus \mathcal{I}_{\nu^G}, \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

$$\hat{\nu}^\Phi := \begin{cases} \frac{-\hat{\nu}_i^{sG}}{\alpha_i^H(\hat{s}; t)} & \text{for } i \in \mathcal{I}_{-0}^{00}(x^*, \lambda), \\ \frac{-\hat{\nu}_i^{sH}}{\alpha_i^G(\hat{s}; t)} & \text{for } i \in \mathcal{I}_{0-}^{00}(x^*, \lambda) \cup \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Once again we run through the 6 conditions. It is to be noted that the variables involved in I. have not been changed so this condition stands true.

II. As pointed out earlier, the gradient of the Lagrangian with respect to s is given by

$$\begin{aligned} \nabla_{s_G} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}^{sG} - \hat{\nu}^G + \hat{\nu}^\Phi \alpha^H(\hat{s}; t), \\ \nabla_{s_H} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}^{sH} - \hat{\nu}^H + \hat{\nu}^\Phi \alpha^G(\hat{s}; t). \end{aligned}$$

For indices i in $\mathcal{I}_{-0}^{00}(x^*, \lambda)$ and $\mathcal{I}_{0-}^{00}(x^*, \lambda)$ we refer to case **(A)**. Let us consider indices i in $\mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}$ and $\mathcal{I}_{\nu^H} \setminus \mathcal{I}_{\nu^G}$. For $i \in \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}$, then $\hat{\nu}_i^G > 0$, $\hat{\nu}_i^H = 0$ and $\hat{\nu}_i^\Phi = \frac{-\hat{\nu}_i^{sH}}{\alpha_i^G(\hat{s}; t)}$ and so the gradient of the Lagrangian with respect to s becomes

$$\begin{aligned} \nabla_{s_{G,i}} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}_i^{sG} - \hat{\nu}_i^G + \hat{\nu}_i^\Phi \alpha_i^H(\hat{s}; t) = \frac{-\hat{\nu}_i^{sH} \alpha_i^H(\hat{s}; t)}{\alpha_i^G(\hat{s}; t)}, \\ \nabla_{s_{H,i}} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) &= \hat{\nu}_i^{sH} - \hat{\nu}_i^H + \hat{\nu}_i^\Phi \alpha_i^G(\hat{s}; t) = 0. \end{aligned}$$

By construction of $\hat{s}_{G,i}$ and $\hat{s}_{H,i}$, it holds that $F_{H_i}(\hat{s}; t) = 0$ and so

$$\nabla_{s_G} \mathcal{L}_s^1(\hat{x}, \hat{s}, \hat{\nu}; t) = \frac{-\hat{\nu}_i^{sH} \alpha_i^H(\hat{s}; t)}{\alpha_i^G(\hat{s}; t)} = \hat{\nu}_i^{sH} \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=s_{G,i}} = \hat{\nu}_i^{sH} \frac{\partial \psi(x; t)}{\partial x} \Big|_{x=\frac{\epsilon}{\hat{\nu}_i^G} - \bar{t}} \leq \epsilon,$$

for some $t \in (0, t^*)$ according to Lemma 9.3.

Now, for indices $i \in \mathcal{I}_{\nu^H} \setminus \mathcal{I}_{\nu^G}$ it holds that $\hat{\nu}_i^{sG} = 0$ so by the choice of multipliers ν^G, ν^H and ν^Φ the gradient of the Lagrangian with respect to s vanishes.

This allows to conclude that the condition II holds true.

III. Since \hat{x} is feasible for the MPCC therefore

$$\begin{aligned} |G_i(\hat{x}) - \hat{s}_{G,i}| &= |\hat{s}_{G,i}| \text{ and } |H_i(\hat{x}) - \hat{s}_{H,i}| = 0, \text{ for } i \in \mathcal{I}^{0+}(x^*), \\ |G_i(\hat{x}) - \hat{s}_{G,i}| &= 0 \text{ and } |H_i(\hat{x}) - \hat{s}_{H,i}| = |\hat{s}_{H,i}|, \text{ for } i \in \mathcal{I}^{+0}(x^*), \\ |G_i(\hat{x}) - \hat{s}_{G,i}| &= |\hat{s}_{G,i}| \text{ and } |H_i(\hat{x}) - \hat{s}_{H,i}| = |\hat{s}_{H,i}|, \text{ for } i \in \mathcal{I}^{00}(x^*). \end{aligned}$$

Let us consider the case where $i \in \mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}$ noticing that $i \in \mathcal{I}_{\nu^H} \setminus \mathcal{I}_{\nu^G}$ is similar by symmetry. The others cases have been checked in case **(A)** of this proof. It follows that

$$\begin{aligned} |\hat{s}_{G,i}| &= \left| \frac{\epsilon - \bar{t} \hat{\nu}_i^G}{\hat{\nu}_i^G} \right| \leq \left| \frac{\epsilon}{\hat{\nu}_i^G} \right| + \bar{t} \leq \bar{t} + c\epsilon, \\ |\hat{s}_{H,i}| &= \left| \psi \left(\frac{\epsilon - \bar{t} \hat{\nu}_i^G}{\hat{\nu}_i^G}; t \right) \right|. \end{aligned}$$

The second condition is ensured by Lemma 9.4.

IV. and **V.** These conditions are straightforward and follows the same path that condition IV. in the case **A**.

VI. The proof for indices is very similar to the condition VI in case **(A)** except in the case where i is in $\mathcal{I}_{\nu^G} \cap \mathcal{I}_{\nu^H}$. In this case

$$\hat{\nu}_i^\Phi = \frac{-\hat{\nu}_i^{s_H}}{\alpha_i^G(\hat{s}; t)} = \frac{-\hat{\nu}_i^{s_H}}{F_{G_i}(\hat{s}; t)},$$

since by construction of \hat{s}_G and \hat{s}_H , $F_{H_i}(\hat{s}; t) = 0$ here. Now, by definition of $F_G(\hat{s}; t)$ it follows

$$\hat{\nu}_i^\Phi = \frac{-\hat{\nu}_i^{s_H}}{\hat{s}_{G,i} - \psi(\hat{s}_{H,i}; t)} = \frac{-\hat{\nu}_i^{s_H}}{\epsilon / \hat{\nu}_i^{s_G} - \bar{t} - \psi(\hat{s}_{H,i}; t)} \geq 0,$$

for $c\epsilon \leq \bar{t}$ since ψ is non-negative whenever s_G and s_H are chosen such that $F_{H_i}(s; t) = 0$ (Assumption (H5)).

The verification of all 6 cases proves that the point constructed above is strong epsilon-stationary, which concludes the proof of the relaxations **(B)** and complete the whole proof. \square

It is to be noted that no constraint qualification is required for this result. This is a clear improvement over what was obtained in the literature in the ideal case of sequence of stationary points. For instance, Theorem 5.1 in [20] requires some second-order information to get a result on existence of stationary points for the regularized sub-problems.

Theorem 9.2. *For any M -stationary point of (MPCC) that satisfies MPCC-CRSC, there exists a sequence of strong epsilon-stationary points of the relaxation $(R_i^s(x, s))$ that converges to that point.*

Proof. Theorem 9.1 gives more relations between the parameters that are compatible with Corollary 8.1. Indeed for a chosen sequence of arbitrarily small parameters $\{\epsilon^k\}$, Corollary 8.1 requires that $\epsilon^k = o(\bar{t}_k)$ and Theorem 9.1 requires that $\bar{t}_k > c\epsilon$ and t_k must be sufficiently small, in particular smaller than $\bar{t}_k - c\epsilon_k$.

Thus, a straightforward application of both of these results provides the result. \square

Previous section point out that such result can not be obtained without a formulation with slack variables.

10 How to Compute Strong Epsilon-Stationary Points

The previous section introduces the new concept of strong epsilon-stationary point of the relaxed sub-problems. In this section, we answer the non-trivial question of how to compute such an approximate stationary point. We present here a generalization of the penalization with active set scheme proposed in [21] and illustrate the fact that it has the desired property.

10.1 A Penalization Formulation

The following minimization problem aims at finding $(x, s) \in \mathbb{R}^n \times \mathbb{R}^{2q}$ so that

$$\begin{aligned} \min_{x,s} \Psi_\rho(x, s) &:= f(x) + \frac{1}{2\rho} \phi(x, s) && (P_\rho^t(x, s)) \\ \text{s.t. } s_G &\geq -\bar{t}\mathbf{1}, \quad s_H \geq -\bar{t}\mathbf{1}, \quad \Phi(s_G, s_H; t) \leq 0, \end{aligned}$$

where ϕ is the penalty function

$$\phi(x, s) := \|\max(g(x), 0), h(x), G(x) - s_G, H(x) - s_H\|^2.$$

An adaptation of Theorem 6.1 gives the following result that validate the penalization approach.

Theorem 10.1. *Given a decreasing sequence $\{\rho_k\}$ of positive parameter and $\{\epsilon_k\}$ a sequence of non-negative parameter that decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \nu^k\}$ be a sequence of strong epsilon-stationary points of $(P_\rho^t(x, s))$ according to Definition 6.1 for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-CRSC holds at x^* . If x^* is feasible, then it is an M -stationary point of (MPCC).*

Proof. Assuming that x^* is feasible for (MPCC), the result is a straightforward adaptation of Theorem 6.1. \square

Unfortunately, the strong assumption on the previous theorem that x^* must be feasible is hard to avoid. Indeed, it is a classical pitfall of penalization methods in optimization to possibly compute a limit point that minimizes the linear combination of the constraints. In other words, we compute a point x^* infeasible that satisfies

$$\sum_{i=1}^p \max(-g_i(x^*), 0) \nabla g_i(x^*) + \sum_{i=1}^m h_i(x^*) \nabla h_i(x^*) - \sum_{i=1}^q \max(G_i(x^*), 0) \nabla G_i(x^*) - \sum_{i=1}^q \max(H_i(x^*), 0) \nabla H_i(x^*) = 0.$$

This phenomenon has been well-known in non-linear programming methods for instance with filter methods. Such a point is sometimes called infeasible stationary point.

It is interesting to note that the way the penalty parameter ρ behave may provide some informations on the stationarity of the limit point. Indeed, if we find a stationary point of the initial problem without driving ρ to zero, then we get an S-stationary point. This observation was introduced in [6] in the context of elastic interior-point for (MPCC) and then adapted to the penalization technique from [21].

Theorem 10.2. *Let (x, s) be a strong epsilon-stationary point of $(P_\rho^t(x, s))$ with $\rho > 0$. If x is feasible for (MPCC), then x is an S-stationary point of (MPCC).*

This fact was already observed in Theorem 2 of [21] in a slightly weaker but similar framework. We do not repeat the proof, but gives an interpretation of this result.

It has been made clear in the proof of the convergence theorem, Theorem 6.1, that the case where x^* is an M-stationary point only occur if the sequence of multipliers $\{\nu^{\Phi, k}\}$ diverges. Therefore, it is to be expected that the penalty parameter must be driven to its limit to observe such phenomenon.

10.2 Overview of the Algorithm

We present here the general steps of our algorithm to compute an M-stationary point of the (MPCC). The algorithm is composed of two loops: the regularization loop, and the penalization-active set loop.

The main loop presented in Algorithm 2 deals with the regularization method. Based on a predefined sequence of precision ϵ_k , we compute the parameters of the relaxation t_k and \bar{t}_k as well as a safeguard parameter used $\rho_{min, k}$ for the penalization. For each of these parameters, we compute a sequence $\{z^k\}$ of iterates and a sequence $\{\rho^k\}$ of penalty parameters using the penalization-active set strategy, such that z^{k+1} is an approximate stationary point of $(P_\rho^t(x, s))$ with $\rho^{k+1} \geq \rho_{min, k}$.

The sequence of iterate is computed through the inner loop dealing with the penalization-active set strategy described in Algorithm 1. The penalization parameter is updated during the process each time not enough progress in term of feasibility is made. However, this reduction can be done only a finite number of time, since the parameter must be larger than the safeguard value $\rho_{min, k}$. Then, Lemma 10.1 and Theorem 10.3 will show that the set of active constraints is also changed a finite number of times and guarantee convergence under mild assumption to a strong-epsilon stationary point, which should be sufficient to get convergence of the whole process according to Theorem 10.1.

In the sequel, we get into the details of the algorithm and its theoretical properties.

10.3 Active Set Method for the Penalized Problem

We discuss here an active set method to solve the penalized problem $(P_\rho^t(x, s))$. This method is an extension of the method proposed in [21] to the general class of methods presented in previous sections.

The set of points that satisfies the constraints of $(R_t^s(x, s))$ is denoted by $\mathcal{F}_{t, \bar{t}}$ and let $\beta_{t, \bar{t}}(x, s)$ denotes the measure of feasibility

$$\begin{aligned} \beta_{t, \bar{t}}(x, s) := & \|\max(g(x), 0)\|^2 + \|h(x)\|^2 + \|G(x) - s_G\|^2 + \|H(x) - s_H\|^2 + \\ & \|\max(-s_G + \bar{t}\mathbf{1}, 0)\|^2 + \|\max(-s_H + \bar{t}\mathbf{1}, 0)\|^2 + \|\max(-\Phi(s_G, s_H; t), 0)\|^2. \end{aligned}$$

Let $\mathcal{W}(s; t, \bar{t})$ be the set of active constraints among the constraints

$$s_G \geq -\bar{t}\mathbf{1}, \quad s_H \geq -\bar{t}\mathbf{1}, \quad \Phi(s_G, s_H; t) \leq 0, \quad (18)$$

and $\mathcal{F}_{t, \bar{t}}^R$ denotes the set of points that satisfies those constraints. We can be even more specific when for some $i \in \{1, \dots, q\}$ the relaxed constraint is active since

$$\Phi_i(s_G, s_H; t) = 0 \implies s_{H,i} = \psi(s_{G,i}; t) \text{ or } s_{G,i} = \psi(s_{H,i}; t).$$

Remark 10.1. *It is essential to note here that active constraints act almost like bound constraints since an active constraint means that for some $i \in \{1, \dots, q\}$ one (possibly both) of the two cases holds*

$$s_{G,i} = -\bar{t} \text{ or } \psi(s_{H,i}; t), \text{ or } s_{H,i} = -\bar{t} \text{ or } \psi(s_{G,i}; t).$$

Considering the relaxation from Kanzow & Schwartz it is obviously a bound constraint since $\psi(s_{G,i}; t) = \psi(s_{H,i}; t) = t$. The butterfly relaxation gives $\psi(s_{G,i}; t) = t_1\theta_{t_2}(s_{H,i})$ and $\psi(s_{H,i}; t) = t_1\theta_{t_2}(s_{G,i})$. This is not a bound constraint but we can easily use a substitution technique. This key observation is another motivation to use a formulation with slack variables.

Furthermore, a careful choice of the function ψ may allow to get an analytical solution of the following equation in α for given values of $s_G, s_H, d_{s_G}, d_{s_H}$:

$$s_{G,i} + \alpha d_{s_{G,i}} - \psi(s_{H,i} + \alpha d_{s_{H,i}}; t) = 0.$$

Solving exactly this equation is very useful while computing the largest step so that the iterates remain feasible along a given direction. For the butterfly relaxation with $\theta_{t_2}(x) = \frac{x}{x+t_2}$, the equation above is reduced to the following second order polynomial equation if $s_{H,i} + \alpha d_{s_{H,i}} \geq 0$:

$$(s_{H,i} + \alpha d_{s_{H,i}} + t_2)(s_{G,i} + \alpha d_{s_{G,i}}) - t_1(s_{H,i} + \alpha d_{s_{H,i}}) = 0. \quad (19)$$

Algorithm 1 presents an active-set scheme to solve $(P_\rho^t(x, s))$, which is described in depth in the sequel of this section. Apart from some specific parameters most of the input data are given in this algorithm through

the relaxation loop that will be discussed in Algorithm 2 (page 35).

	Data:
	Input Data: x^{k-1}, s^{k-1} ; precision $\epsilon > 0$;
	$\rho_0 > 0$ initial value of ρ , ρ_{min} lower bound on the penalty parameter ;
	Algorithm Parameters: $\sigma_\rho \in (0, 1)$ update in ρ ; $\tau_{vio} \in (0, 1)$; sat:=true;
	Initial estimate of the multiplier ν^0 ;
1	Begin ;
2	Set $j := 0, \rho := \rho_0$;
3	$(x^{k,0}, s^{k,0}, \mathcal{W}_0, A_0)$ =Projection of (x^{k-1}, s^{k-1}) if not feasible for $(P_\rho^t(x, s))$;
4	while sat and $(\ \nabla \mathcal{L}^1(x^{k,j}, s_G^{k,j}, s_H^{k,j}, \nu^j; t_k)\ _\infty^2 > \epsilon \ \nu^j\ _\infty$ or $\min(\nu^j) < 0$ or $\beta_{t_k, \bar{t}_k}(x^{k,j}, s^{k,j}) > \epsilon)$
	do
5	Substitution of the variables that are fixed by the active constraints in \mathcal{W}_j ;
6	Compute a feasible direction d^j that lies in the subspace defined by the working set \mathcal{W}_j (see (21)) and satisfies the conditions (SDD);
7	Compute $\bar{\alpha}$ the maximum non-negative feasible step along d^j
	$\bar{\alpha} := \sup\{\alpha : (x^{k,j}, s^{k,j}) + \alpha d^j \in \mathcal{F}_{t, \bar{t}}\}$
	Compute a step length $\alpha_j \leq \bar{\alpha}$ (see (19)) such that Armijo condition (22) holds ;
8	if $\alpha_j = \bar{\alpha}$ then
9	Update the working set $\rightarrow \mathcal{W}_{j+1}$ and compute A_{j+1} the matrix of gradients of active constraints
10	$(x^{k,j+1}, s^{k,j+1}) = (x^{k,j}, s^{k,j}) + \alpha_j d^j$;
11	$j:=j+1$;
12	if $\beta_{t_k, \bar{t}_k}(x^{k,j+1}, s^{k,j+1}) \geq \max(\tau_{vio} \beta_{t_k, \bar{t}_k}(x^{k,j}, s^{k,j}), \epsilon)$ then
13	$\rho := \max(\sigma_\rho \rho, \rho_{min})$
14	else
15	Determine the approximate multipliers $\nu^{j+1} = (\nu^G, \nu^H, \nu^\Phi)$ by solving
	$\nu^{j+1} \in \arg \min_{\nu \in \mathbb{R}^{ \mathcal{W}_j }} \ A_{j+1}^T \nu - \nabla \Psi_\rho(x^{k,j}, s^{k,j})\ ^2$
	Relaxing rule : if $\exists i, \nu_i^{j+1} < 0$ and (satisfy (23) or $\alpha_j = 0$) then
16	Update of the working set \mathcal{W}_{j+1} (with an anti-cycling rule) ;
17	sat:= $\ d^j\ > \epsilon$
18	return: x^k, s^k, ρ or a decision of unboundedness.

Algorithm 1: Active-Set Penalization Algorithm for the relaxed non-linear program $(P_\rho^t(x, s))$.

At each step, the set \mathcal{W}_j denotes the set of active constraints of the current iterate $s^{k,j}$. As pointed out in Remark 10.1, these active constraints fix some of the variables. Therefore, by replacing these fixed variables we can rewrite the problem in a subspace of the initial domain. Thus, we consider the following minimization problem

$$\begin{aligned}
 & \min_{(x,s) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{S}_G|+|\mathcal{S}_H|}} \Psi_\rho(x, s_{\mathcal{S}_G \cup \mathcal{S}_H}) \\
 & \text{s.t. } s_{G,i} \geq -\bar{t} \text{ for } i \in \mathcal{S}_G, \quad s_{H,i} \geq -\bar{t} \text{ for } i \in \mathcal{S}_H, \\
 & \quad \Phi_i(s_G, s_H; t) \leq 0 \text{ for } i \in \mathcal{S}_G \cup \mathcal{S}_H,
 \end{aligned} \tag{20}$$

where we denote

$$\begin{aligned}
\mathcal{I}_G &:= \{i \in \{1, \dots, q\} \mid s_{Gi} = -\bar{t}\}, \\
\mathcal{I}_H &:= \{i \in \{1, \dots, q\} \mid s_{Hi} = -\bar{t}\}, \\
\mathcal{I}_{GH}^{0+} &:= \{i \in \{1, \dots, q\} \mid s_{Hi} = \psi(s_G; t)\}, \\
\mathcal{I}_{GH}^{+0} &:= \{i \in \{1, \dots, q\} \mid s_{Gi} = \psi(s_H; t)\}, \\
\mathcal{I}_{GH}^{00} &:= \{i \in \{1, \dots, q\} \mid s_{Gi} = s_{Hi} = \psi(0; t)\}, \\
\mathcal{S}_G &:= \{i \in \{1, \dots, q\}\} \setminus (\mathcal{I}_G \cup \mathcal{I}_{GH}^{+0} \cup \mathcal{I}_{GH}^{00}), \\
\mathcal{S}_H &:= \{i \in \{1, \dots, q\}\} \setminus (\mathcal{I}_H \cup \mathcal{I}_{GH}^{0+} \cup \mathcal{I}_{GH}^{00}).
\end{aligned}$$

\mathcal{S}_G and \mathcal{S}_H respectively denote the set of indices where the variables s_G and s_H are free.

Some of the fixed variables are replaced by a constant and others are replaced by an expression that depends on the free variables. It is rather clear from this observation that the use of slack variables is an essential tool to handle the non-linear bounds.

A major consequence here is that the gradient of Ψ in this subspace can be done using the composition of the derivative formula:

$$\nabla \Psi_\rho^{\mathcal{W}_j}(x, s_{\mathcal{S}_G \cup \mathcal{S}_H}) = J_{\mathcal{W}_j}^T \nabla \Psi_\rho(x, s), \quad (21)$$

where $J_{\mathcal{W}_j}$ is an $(n + 2q) \times (n + \#\mathcal{S}_G + \#\mathcal{S}_H)$ matrix defined such that

$$J_{\mathcal{W}_j} := \begin{pmatrix} J_{\mathcal{W}_j}^x \\ J_{\mathcal{W}_j}^{s_G} \\ J_{\mathcal{W}_j}^{s_H} \end{pmatrix}.$$

The three sub-matrices used to define $J_{\mathcal{W}_j}$ are computed in the following way

$$\begin{aligned}
J_{\mathcal{W}_j}^x &= Id_n, \\
J_{\mathcal{W}_j}^{s_G} &= \begin{cases} e_i^T, & \text{for } i \in \mathcal{S}_G, \\ \frac{\partial \psi(x;t)}{\partial x} \Big|_{x=s_H} e_i^T, & \text{for } i \in \mathcal{I}_{GH}^{+0}, \\ 0, & \text{for } i \in (\{1, \dots, q\} \setminus \mathcal{S}_G) \setminus \mathcal{I}_{GH}^{+0}, \end{cases} \\
J_{\mathcal{W}_j}^{s_H} &= \begin{cases} e_i^T, & \text{for } i \in \mathcal{S}_H, \\ \frac{\partial \psi(x;t)}{\partial x} \Big|_{x=s_G} e_i^T, & \text{for } i \in \mathcal{I}_{GH}^{0+}, \\ 0, & \text{for } i \in (\{1, \dots, q\} \setminus \mathcal{S}_H) \setminus \mathcal{I}_{GH}^{0+}, \end{cases}
\end{aligned}$$

where $J_{\mathcal{W}_j, i}$ denotes the i -th line of a matrix and e_i is a vector of zero whose i -th component is one. We may proceed in a similar way to compute the hessian matrix of $\Psi_\rho(x, s_{\mathcal{S}_G \cup \mathcal{S}_H})$.

The feasible direction d^j is constructed to lie in a subspace defined by the working set and satisfying the sufficient-descent direction conditions for $z^j \in \mathbb{R}^{n+|\mathcal{S}_G|+|\mathcal{S}_H|}$:

$$\begin{aligned}
\nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j &\leq -\mu_0 \|\nabla \Psi^{\mathcal{W}_j}(z^j)\|^2, \\
\|d^j\| &\leq \mu_1 \|\nabla \Psi^{\mathcal{W}_j}(z^j)\|,
\end{aligned} \quad (\text{SDD})$$

where $\mu_0 > 0$, $\mu_1 > 0$.

The step length $\alpha_j \in (0, \bar{\alpha}]$ is respectively computed to satisfy the Armijo and Wolfe conditions for $z^j \in \mathbb{R}^{n+|\mathcal{S}_G|+|\mathcal{S}_H|}$:

$$\Psi(z^j + \alpha_j d^j) \leq \Psi(z^j) + \tau_0 \alpha_j \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j, \quad \tau_0 \in (0, 1), \quad (22)$$

$$\nabla \Psi^{\mathcal{W}_j}(z^j + \alpha_j d^j)^T d^j \geq \tau_1 \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j, \quad \tau_1 \in (\tau_0, 1). \quad (23)$$

If $\bar{\alpha}$ satisfies the Armijo condition (22), the active set strategy adds a new active constraint and the Wolfe condition (23) is not enforced. Otherwise, the Armijo condition requires $\alpha < \bar{\alpha}$ and the Wolfe condition is enforced.

The relaxing rule is given by the following scheme : Relax some constraint i_0 if and only if the two following conditions are fulfilled:

1. $\nu_{i_0}^j < 0$;
2. No constraint was added at the arrival point $(x^{k,j}, s^{k,j})$ and no constraint was deleted at the previous iteration.

The convergence result will rely on the fact that at least one step satisfying Wolfe's condition will be performed before removing an active constraint.

Remark 10.2. *One should pay attention here that the algorithm is handling simultaneously the active-set strategy as well as the penalization, since the penalty parameter can be reduced if the feasibility is not reduced enough in step 13. This reduction can only happen a finite number of time ($\rho \geq \rho_{min}$), which will be of importance to show the convergence of the algorithm. Since we expect to compute a strong ϵ -stationary, ρ_{min} depends on ϵ .*

10.4 Convergence of Algorithm 1

We now present an extension of the convergence analysis of [21] to Algorithm 1.

We make the following additional assumption on the problem:

- f is bounded from below;
- the gradients of the objective and constraints satisfy a Lipschitz condition.

Therefore, Ψ is also bounded from below, and $\nabla\Psi$ is Lipschitz with a constant denoted L_∇ .

We restrict our study to the class of methods from the UF Framework that satisfies the following assumption

$$\Phi_i(G(x), H(x); t) = 0 \iff F_{G_i}(x; t) = 0 \text{ or } F_{H_i}(x; t) = 0. \quad (24)$$

This is a stronger assumption than (H4) presented earlier (page 8). This assumption excludes the relaxation KS. However, it still includes a large range of methods including the method KDB introduced earlier, but also,

- the approximate butterfly

$$\Phi(a, b) = (b - t_1\theta_{t_2}(a)) (a - t_1\theta_{t_2}(b));$$

- the modified butterfly

$$\Phi(a, b) = \begin{cases} < 0, & \text{if } F_{1_i}(x; t) < 0 \ \& \ F_{2_i}(x; t) < 0, \\ (b - t_1\theta_{t_2}(a)) (a - t_1\theta_{t_2}(b)), & \text{otherwise.} \end{cases}$$

As pointed out in previous studies on active-set methods in the literature, the most important aspect of the convergence is the relaxing rule. So that, the working set should not be reduced unless the relaxing rule is satisfied. However, it is not clear for us, to know how to react in the situation, where $s_{G,i} = \psi(s_{H,i}; t)$ (or $s_{H,i} = \psi(s_{G,i}; t)$) and $\Phi_i(G(x), H(x); t) < 0$.

We now move to the proof of the convergence of Algorithm 1. Thanks to the several remarks made in the previous section, we can adapt the proof given in [21] by following the same path.

Lemma 10.1. Let $\{z^j\} \in \mathbb{R}^{n+|S_G|+|S_H|}$ be a sequence computed on some active set \mathcal{W} by using the relaxing rule. For each $j \in \mathbb{N}$, it holds that

$$z^{j+1} = z^j + \alpha_j d_j,$$

where d_j satisfies (SDD) and α_j satisfies (22), (23). Let $\{\rho_j\}$ be a sequence computed according to step 13 in Algorithm 1. Then, for any converging subsequence $J \subset \mathbb{N}$, it holds that

$$\forall \epsilon > 0, \exists \bar{j} \in \mathbb{N} \implies \forall j \geq \bar{j}, x^j \text{ is an } \epsilon\text{-stationary point of (20) with } \rho_j \geq \rho_{min}.$$

Proof. Consider any working set \mathcal{W} visited infinitely often but possessing a finite number of pairs of consecutive iterates. For such a working set, $\bar{\alpha}$ is used infinitely often and cluster points actually belong to some other working set with at least one more active constraint than \mathcal{W} . It suffices then to consider that the sequence $\{z^j\}$ possesses infinitely many pairs of consecutive iterates on \mathcal{W} .

Furthermore, the update of the penalty parameter is done only a finite number of time, since it is bounded below by ρ_{min} and reduced at each update step. Thus, the sequence $\{\rho_j\}$ converges to some value ρ greater or equal than ρ_{min} in a finite number of iteration.

By conditions (SDD), we have $\Psi(z^{j+1}) \leq \Psi(z^j)$, for all j . By hypothesis, for all $\bar{\epsilon} \geq 0$, there exists a \bar{j} such that for all $J \ni j \geq \bar{j}$

$$\|\Psi(z^{j+1}) - \Psi(z^j)\| \leq \bar{\epsilon}. \quad (25)$$

Armijo condition, (22), is satisfied by every iterate,

$$\Psi(z^{j+1}) \leq \Psi(z^j) + \tau_0 \alpha_j \nabla \Psi^{\mathcal{W}}(z^j)^T d^j.$$

Since $\nabla \Psi^{\mathcal{W}}(z^j)^T d^j \leq 0$, we have

$$\begin{aligned} \Psi(z^{j+1}) &\leq \Psi(z^j) + \tau_0 \alpha_j \nabla \Psi^{\mathcal{W}}(z^j)^T d^j \leq \Psi(z^j), \\ \iff \Psi(z^{j+1}) - \Psi(z^j) &\leq \tau_0 \alpha_j \nabla \Psi^{\mathcal{W}}(z^j)^T d^j \leq 0. \end{aligned}$$

Now, for all $j \geq \bar{j}$, (25) and previous inequalities give

$$\|\alpha_j \nabla \Psi^{\mathcal{W}}(z^j)^T d^j\| \leq \frac{\bar{\epsilon}}{\tau_0}. \quad (26)$$

Suppose that z^j is not a stationary point of the restricted problem (20). Then, there exists $M > 0$ such that for j sufficiently large, we have:

$$\|\nabla \Psi^{\mathcal{W}}(z^j)\| \geq M. \quad (27)$$

By using the first condition of (SDD), we have

$$\alpha_j \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j \leq -\mu_0 \alpha_j \|\nabla \Psi^{\mathcal{W}_j}(z^j)\|^2 < 0.$$

Using assumption (27), we get

$$\alpha_j \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j \leq -\mu_0 \alpha_j M^2.$$

Using (26), we obtain

$$\|\alpha_j\| \leq \frac{\bar{\epsilon}}{\tau_0 \mu_0 M^2}. \quad (28)$$

The second part of (SDD) gives

$$\|\alpha_j d^j\| \leq \mu_1 \alpha_j \|\nabla \Psi^{\mathcal{W}_j}(z^j)\| \leq \mu_1 \alpha_j L_{\nabla} \|z^j\|,$$

using Lipschitz assumption. Now, by (28), we have

$$\|\alpha_j d^j\| \leq \bar{\epsilon} \frac{\mu_1 L_{\nabla} \|z^j\|}{\tau_0 \mu_0 M^2}. \quad (29)$$

Wolfe condition, (23), gives that

$$\nabla\Psi^{\mathcal{W}_j}(z^j + \alpha_j d^j)^T d^j \geq \tau_1 \nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j,$$

and subtracting $\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j$ yields to

$$\|(\nabla\Psi^{\mathcal{W}_j}(z^j + \alpha_j d^j) - \nabla\Psi^{\mathcal{W}_j}(z^j))^T d^j\| \geq (1 - \tau_1) \|\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j\|.$$

Using Cauchy-Schwarz inequality, we obtain

$$\|\nabla\Psi^{\mathcal{W}_j}(z^j + \alpha_j d^j) - \nabla\Psi^{\mathcal{W}_j}(z^j)\| \|d^j\| \geq (1 - \tau_1) \|\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j\|.$$

By Lipschitz assumption, it follows that

$$\alpha_j L_{\nabla} \|d^j\|^2 \geq (1 - \tau_1) \|\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j\|.$$

Using the second condition in (SDD) and (29) leads to

$$\|\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j\| \leq \bar{\epsilon} \frac{\mu_1^2 L_{\nabla}^3 \|z^j\|}{\tau_0 \mu_0 M^2 (1 - \tau_1)}. \quad (30)$$

Using the first condition in (SDD) gives

$$\|\nabla\Psi^{\mathcal{W}_j}(z^j)^T d^j\| \geq \mu_0 \|\nabla\Psi^{\mathcal{W}_j}(z^j)\|^2.$$

By equation (30) and (27), we have

$$\bar{\epsilon} \frac{\mu_1^2 L_{\nabla}^3 \|z^j\|}{\tau_0 \mu_0^2 M^2 (1 - \tau_1)} \geq M^2. \quad (31)$$

This, however, leads to a contradiction, since $\bar{\epsilon}$ can be made arbitrarily small with the only consequence that \bar{j} becomes larger. Here, we use that since J is a converging subsequence, there exists a positive constant that bounds the sequence.

Thus, (27) can not hold, which proves the result. \square

In order to prove the global convergence of Algorithm 1 we enrich the relaxation conditions by assuming that the current iterate satisfied the Wolfe conditions. This is denoted as *sufficient-long relaxing step* in [21]. We now state the result, which once again is a straightforward application of Theorem 3 from [21].

Theorem 10.3. *Let $\{z^j\} \in \mathbb{R}^{n+2q}$ and $\{\rho_j\}$ be sequences computed by Algorithm 1. Then, for any converging subsequence $J \subset \mathbb{N}$, it holds that*

$$\forall \epsilon > 0, \exists \bar{j} \implies \forall j \geq \bar{j}, x^j \text{ is a strong } \epsilon\text{-stationary point of } (P_{\rho}^t(x, s)) \text{ with } \rho_j \geq \rho_{min}.$$

Proof. First, we verify the result for (not necessarily strong) ϵ -stationary point of $(P_{\rho}^t(x, s))$.

Since, from Lemma 10.1, for each subsequence $J \subset \mathbb{N}$, up to some rank, the iterates are ϵ -stationary points of some restricted problem, it suffices to examine an infinite subsequence $\{z^j\}$ of relaxing steps.

In the same way as in Lemma 10.1, the sequence of penalty parameter $\{\rho_j\}$ converges in finite number of iteration to some value ρ greater or equal to ρ_{min} .

Let z^{j+} denote the point reached after a relaxation step has been taken

$$\Psi(z^{j+}) \leq \Psi(z^j) + \tau_0 \alpha_j \nabla\Psi^{\mathcal{W}}(z^j)^T d_j.$$

Since d^{j+} is a descent direction then

$$\Psi(z^{j+1}) \leq \Psi(z^{j+}) \leq \Psi(z^j) + \tau_0 \alpha_j \nabla\Psi^{\mathcal{W}}(z^j)^T d_j. \quad (32)$$

Let $\epsilon \geq 0$ be fixed. Assume that there is no \bar{j} such that for all $j \geq \bar{j}$, z^j is an ϵ -stationary point of $(P_\rho^t(x, s))$. The relaxing rule force the algorithm to take a relaxing step and the *sufficient-long relaxing step*, (23), implies for $j \in [\bar{j}, \infty]$

$$\alpha_j > \delta, \quad (33)$$

with δ independent of j . By using the sufficient-descent direction property of d^j , for j sufficiently large, we have

$$\alpha_j \nabla \Psi^{\mathcal{W}}(z^j)^T d_j \leq -\mu_0 \alpha_j \|\nabla \Psi^{\mathcal{W}}(z^j)\|^2 \leq -\mu_0 \alpha_j K,$$

for some $K > 0$ (non-optimality implies $\|\nabla \Psi^{\mathcal{W}}(z^j)\| > \epsilon$). Then, (32) implies

$$\Psi(z^{j+1}) \leq \Psi(z^j) + \tau_0 - \mu_0 \alpha_j K.$$

Now, since (33), we have

$$\lim_{j \rightarrow \infty} \Psi(z^j) = -\infty,$$

which contradicts the fact that Ψ is bounded below, since f is bounded below. This concludes that necessarily there exists an index \bar{j} such that for all $j \geq \bar{j}$, z^j is an ϵ -stationary point of $(P_\rho^t(x, s))$.

The strong ϵ -stationary property trivially follows, since each iterates satisfy exactly the relaxed complementarity constraint, by the active-set strategy, and also the complementarity condition, since the support of the multipliers are computed exactly. \square

10.5 An Algorithm for MPCC

Along this paper, we analyze an algorithm to solve (MPCC) through a regularization scheme and an active set-penalization method to solve the sub-problems. The latter has been described in the previous sections. We now formally defined the regularization scheme in Algorithm 2.

Data: Let $z^0 = (x^0, s^0)$ be an initial point;
 Let ρ_0 be an initial value of the penalty parameter;
 Choose a sequence of precision $\{\epsilon_k\}$, a desired precision ϵ_∞ and a safeguard ϵ_{min} ;
 Set $k = 0$;
1 Begin ;
2 repeat
3 $(t_k, \bar{t}_k, \rho_{min,k}) := \text{Oracle}(\epsilon_k)$;
4 $z^{k+1}, \rho_{k+1} = \text{Algorithm1}(z^k, \epsilon_k, \rho_k, \rho_{min,k})$: from the starting point z^k , use Algorithm 1 to compute z^{k+1} an approximate stationary point of $(P_\rho^t(x, s))$ with penalty parameter $\rho_k \geq \rho_{k+1} \geq \rho_{min,k}$;
5 Set $k \leftarrow k + 1$;
6 until $(\|\min(G(x^{k+1}), H(x^{k+1}))\|_\infty \leq \epsilon_\infty \text{ and } \phi(z^{k+1}) \leq \epsilon_\infty) \text{ or } \epsilon_k < \epsilon_{min}$;
7 return: $f_{opt} := f(x^{k+1})$ the optimal value at the solution $x_{opt} := x^{k+1}$ or a decision of infeasibility or unboundedness.

Algorithm 2: Relaxation method for Problem (MPCC).

In step 3 of Algorithm 2, an oracle compute (t, \bar{t}, ρ_{min}) using the chosen value of ϵ_k that respect the conditions discussed in Section 9.2, while the value of $\rho_{min,k}$ should also depend on the precision. Indeed, approximate feasibility of the iterates for k sufficiently large is not guaranteed by Algorithm 1, since we can reduce finitely many times ρ to some value ρ_{min} . So, this safeguard value should be chosen sufficiently small.

According to Theorem 10.1 and Theorem 10.3, whenever the final iterate x^k in Algorithm 2 is feasible for (MPCC) and satisfies MPCC-CRSC, then it is an M-stationary point up to some precision $\bar{\epsilon}$.

Definition 10.1. x^* is said an M stationary point up to $\bar{\epsilon}$, if there exists $\lambda \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$ satisfying

$$\begin{aligned} \|\nabla_x \mathcal{L}_{MPCC}^1(x^*, \lambda)\| &\leq \bar{\epsilon}, \\ \|\min(-g(x^*), \lambda^g)\| &\leq \bar{\epsilon}, \quad \|h(x^*)\| \leq \bar{\epsilon}, \quad \|\min(G(x^*), H(x^*))\| \leq \bar{\epsilon}, \\ \|G_i(x^*)\lambda_i^G\| &\leq \bar{\epsilon}, \quad \|H_i(x^*)\lambda_i^H\| \leq \bar{\epsilon}, \quad \|\min(|\lambda_i^G \lambda_i^H|, \max(-\lambda_i^G, 0) + \max(-\lambda_i^H, 0))\| \leq \bar{\epsilon} \quad \forall i \in \{1, \dots, q\}. \end{aligned}$$

Finally, we can conclude the convergence of the whole process.

Theorem 10.4. Given $\{\epsilon_k\}$ a sequence of non-negative parameter that decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, s^k\}$ be a sequence of points computed by Algorithm 2. If for any k sufficiently large $\phi(x^k, s^k) \leq \epsilon_k$ and MPCC-CRSC holds at x^k , then it holds that

$$\forall \epsilon > 0, \exists \bar{k} \implies \forall k \geq \bar{k}, x^k \text{ is an } M\text{-stationary point of (MPCC) up to } \epsilon.$$

Proof. Theorem 10.3 guarantees convergence of Algorithm 1 to a strong ϵ -stationary point of $(P_\rho^t(x, s))$.

Therefore, Algorithm 2 computes a sequence of strong ϵ -stationary point, which converges to an M -stationary point under the stated assumptions as proved by Theorem 10.3. \square

The following corollary is a direct consequence and state the convergence of our approach. This is our main result proving convergence of the regularization-penalization-active set scheme to solve (MPCC).

Corollary 10.1. Given $\{\epsilon_k\}$ a sequence of non-negative parameter that decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^*, s^*\}$ be an accumulation point of the sequence of points $\{x^k, s^k\}$ computed by Algorithm 2. If x^* is feasible for (MPCC) and MPCC-CRSC holds at x^* , then x^* is an M -stationary point.

This algorithm is the first from the literature offering guarantees of convergence for (MPCC). Indeed, we remind here that most of the methods proposed face the problem of dealing with approximate stationary points as pointed out in [23].

11 Numerics

In what follows, we present a small set of instances to show the behaviour of our algorithm. Beforehand, we give some supplementary informations regarding the implementation of Algorithm 2 and Algorithm 1. An extended butterfly relaxation has been used, which consider $t \in \mathbb{R}^3$ and

$$\psi(z; t) = t_3 + t_2 \theta_{t_1}^1(z - t_3),$$

where $\theta_{t_1}^1(z - t_3) = \frac{z - t_3}{z - t_3 + t_1}$ for $z \geq t_3$ and $\theta_{t_1}^1(z - t_3) = \frac{z - t_3}{t_1} - \frac{(z - t_3)^2}{2t_1^2}$ for $z < t_3$. The list of parameters used in the process is detailed in Table 1. It is to be noted that no attempt has been made to optimize the performance of the algorithm and the results come from a straightforward implementation of the algorithm in the JULIA programming language. The direction d^j used in Algorithm 1 is computed through a Newton method. The computation of the constrained step length along this direction is computed through a backtracking line search technique. A comparison between some methods to compute the descent direction has been conducted in [21].

We now introduce two examples and give the result of our method.

The first example is the continuation of Example 5.1, which illustrate a case where $(0, 0)$ is a weak-stationary point.

Example 11.1. Consider the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & x_1 - x_2 \\ \text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0, \quad x_2 \leq 1. \end{aligned}$$

parameter	function	default value
Parameters for Algorithm 2		
t_0	relaxation parameter	(0.1,0.1,0.01)
σ_t	update of relaxation parameter	(0.1,0.1,0.01)
\bar{t}		t_1
ϵ	sequence of precision	$\max(r, s, t)$
ϵ_∞	precision of (MPCC)	10^{-4}
Parameters for Algorithm 1		
ρ_0		0.1
σ_ρ		0.5
ρ_{min}		10^{-8}
τ_{vio}		0.5
τ_0	Armijo parameter	0.25
τ_1	Wolfe parameter	0.9

Table 1: List of parameter for Algorithm 2 and their default values.

Iter	x^k	s^k	$f(x^k)$	ρ^{-1}	# inner Iter
0	(1.0,1.0)	-	0.0	-	-
1	(-0.2,1.1)	(-0.1,1.1)	-1.3	10.0	5
2	(-0.06,1.05)	(-0.06,1.05)	-1.11	20.0	5
3	(-0.026,1.025)	(-0.001,1.025)	-1.051	40.0	4
4	(-0.0126,1.0125)	(-0.0001,1.0)	-1.0251	80.0	4
5	(-0.003135,1.00313)	(-1.0e-5,1.00313)	-1.00626	320.0	5
6	(-0.003126,1.00156)	(-1.0e-6,1.00156)	-1.003126	640.0	5
7	(-0.00078135,1.0)	(-1.0e-7,1.00078)	-1.0015626	1280.0	4
8	(-0.000195322,1.0002)	(-1.0e-8,1.002)	-1.000390635	5120.0	5
9	(-9.76572e-5,1.0001)	(-1.0e-9,1.0001)	-1.0001953135	10240.0	4

Table 2: Solutions and optimal values of Example 11.1.

By starting from the initial point $(x_1, x_2) = (1.0, 1.0)$, our algorithm finds the solution after 9 iterations. Table 2 summarizes the results of the algorithm.

The last example is a continuation of Example 2.1 and illustrates a case where the solution $(0, 0, 0)^t$ is an M-stationary point.

Example 11.2. *Consider the problem*

$$\begin{aligned}
& \min_{x \in \mathbb{R}^3} && x_1 + x_2 - x_3 \\
& s.t. && -4x_1 + x_3 \leq 0, \\
& && -4x_2 + x_3 \leq 0, \\
& && 0 \leq x_1 \perp x_2 \geq 0.
\end{aligned}$$

By starting from the initial point $(x_1, x_2, x_3) = (0.5, 1.0, 1.0)$, our algorithm finds the solution after 8 iterations. Table 3 summarizes the results of the algorithm.

12 Concluding Remarks

A generalized framework presented in this article is used to analyze relaxation methods that aim to converge to M-stationary points. Motivated by the approximate resolution of the sub-problems we defined a new notion

Iter	x^k	s^k	$f(x^k)$	ρ^{-1}	# inner Iter
0	(0.5,1.0,0.0)	-	1.5	-	-
1	(0.207806,0.214056,0.868723)	(0.105328,0.214056)	-0.44686139008	20.0	4
2	(0.0600906,0.0632156,0.259112)	(0.0100842,0.0632156)	-0.13580620018	40.0	4
3	(0.026001,0.0275635,0.113379)	(0.00100096,0.0275635)	-0.05981445899	80.0	4
4	(0.00635001,0.00674063,0.0277438)	(0.00010001,0.00674063)	-0.01465314485	320.0	5
5	(0.0015725,0.00167016,0.00687594)	(1.00001e-5,0.00167016)	-0.00363328144	1280.0	5
6	(0.00078225,0.000831078,0.00342197)	(1.0e-6,0.000831078)	-0.00180864062	2560.0	4
7	(0.000390725,0.000415139,0.00170938)	(1.0e-7,0.000415139)	-0.00090352031	5120.0	4
8	(0.000195323,0.00020753,0.000854532)	(1.0e-8,0.00020753)	-0.00045168015	10240.0	4

Table 3: Solutions and optimal values of Example 11.2.

of approximate stationary point. We proved existence of such approximate point in the neighbourhood of an M-stationary point and provided an algorithmic strategy to compute such point.

The final section provide some preliminary numerical results to validate our approach. Although no attempt has been made to improve the code and we do believe that in many ways it could be improved. Further research concerns the implementation of this algorithmic strategy in JULIA and comparison with existing methods on the library of test problem MacMPEC, [24]. Theoretical analysis of the algorithm proposed here is the subject of future research.

Acknowledgements

This research was partially support by a grant from "l'Ecole des Docteurs de l'UBL" and "le Conseil Régional de Bretagne". This research was partially supported by NSERC grant.

References

- [1] L. Abdallah and M. Haddou. An exact penalty approach for mathematical programs with equilibrium constraints. *Journal of Advances in Mathematics*, 9(8), 2013.
- [2] S. Albrecht and M. Ulbrich. Mathematical programs with complementarity constraints in the context of inverse optimal control for locomotion. *Optimization Methods and Software*, 6788(October):1–29, sep 2016.
- [3] M. Bergounioux and M. Haddou. A new relaxation method for a discrete image restoration problem. *Journal of Convex Analysis*, 17(3):421–435, 2010.
- [4] J.-F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.
- [5] F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5. Siam, 1990.
- [6] Z. Coulibaly and D. Orban. An ℓ_1 Elastic Interior-Point Method for Mathematical Programs with Complementarity Constraints. *SIAM Journal on Optimization*, 22(1):187–211, jan 2012.
- [7] V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes. A two-sided relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 16(2):587–609, 2005.
- [8] S. Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

- [9] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova. *Bilevel programming problems: theory, algorithms and applications to energy networks*. Springer, 2015.
- [10] J.-P. Dussault, M. Haddou, and T. Migot. The new butterfly relaxation methods for mathematical program with complementarity constraint. Technical report, 2016.
- [11] M. L. Flegel. *Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints*. PhD thesis, University of Würzburg, 2005.
- [12] M. L. Flegel and C. Kanzow. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005.
- [13] M. L. Flegel and C. Kanzow. On m-stationary points for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 310(1):286–302, 2005.
- [14] M. L. Flegel and C. Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints*. Springer, 2006.
- [15] L. Guo, G.-H. Lin, and J. J. Ye. Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 158(1):33–64, 2013.
- [16] M. Hintermüller and I. Kopacka. Mathematical Programs with Complementarity Constraints in Function Space: C- and Strong Stationarity and a Path-Following Algorithm. *SIAM Journal on Optimization*, 20(2):868–902, jan 2009.
- [17] T. Hoheisel. *Mathematical programs with vanishing constraints*. PhD thesis, 2009.
- [18] T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1-2):257–288, feb 2013.
- [19] A. Kadrani. *Nouvelles Régularisations pour l’optimisation avec contraintes de complémentarité*. PhD thesis, Université de Sherbrooke, 2008.
- [20] A. Kadrani, J.-P. Dussault, and A. Benchakroun. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 20(1):78–103, 2009.
- [21] A. Kadrani, J.-P. Dussault, and A. Benchakroun. A globally convergent algorithm for mpcc. *EURO Journal on Computational Optimization*, 3(3):263–296, 2015.
- [22] C. Kanzow and A. Schwartz. A New Regularization Method for Mathematical Programs with Complementarity Constraints with Strong Convergence Properties. *SIAM Journal on Optimization*, 23(2):770–798, apr 2013.
- [23] C. Kanzow and A. Schwartz. The Price of Inexactness: Convergence Properties of Relaxation Methods for Mathematical Programs with Complementarity Constraints Revisited. *Mathematics of Operations Research*, 40(2):253–275, may 2015.
- [24] S. Leyffer. Macmpec: Ampl collection of mpecs. *Argonne National Laboratory. Available at www.mcs.anl.gov/leyfier/MacMPEC*, 2000.
- [25] G.-H. Lin and M. Fukushima. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, 133(1-4):63–84, 2005.
- [26] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

- [27] J. V. Outrata. Mathematical Programs with Equilibrium Constraints: Theory and Numerical Methods. In *Nonsmooth Mechanics of Solids*, number 1, pages 221–274. Springer Vienna, Vienna, 2002.
- [28] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.
- [29] S. Scholtes and M. Stöhr. How stringent is the linear independence assumption for mathematical programs with complementarity constraints? *Mathematics of Operations Research*, 26(4):851–863, 2001.
- [30] A. Schwartz. *Mathematical programs with complementarity constraints: Theory, methods, and applications*. PhD thesis, Ph. D. dissertation, Institute of Applied Mathematics and Statistics, University of Würzburg, 2011.
- [31] S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.
- [32] S. Veelken. *A new relaxation scheme for mathematical programs with equilibrium constraints: Theory and numerical experience*. PhD thesis, Julius-Maximilians-Universität Würzburg, 2009.
- [33] J. J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005.
- [34] J. J. Ye and X. Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, 22(4):977–997, 1997.