The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints

Tangi Migot, Mounir Haddou, Jean-Pierre Dussault

The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints

Migot, T* Haddou, M† Dussault, J.-P.‡

2016

Abstract

We propose a new family of relaxation schemes for mathematical programs with complementarity constraints that extends the relaxations of Kadrani, Dussault, Bechakroun from 2009 and the one of Kanzow & Schwartz from 2011. We discuss the properties of the sequence of relaxed non-linear program as well as stationarity properties of limiting points. A sub-family of our relaxation schemes has the desired property of converging to an M-stationary point. We introduce a new constraint qualification, MPCC-CRSC, to prove convergence of our method, which is the weakest known constraint qualification that ensures boundedness of the sequence generated by the method. A comprehensive numerical comparison between existing relaxations methods is performed on the library of test problems MacMPEC and shows promising results for our new method. Numerical perspectives shows an enhanced version of the butterfly relaxation to mathematical program with vanishing constraints.

Keywords : nonlinear programming - MPCC - MPEC - relaxation methods - stationarity - constraint qualification - MPVC - CRSC

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

1 Introduction

We consider the Mathematical Program with Complementarity Constraint

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

with $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $G, H : \mathbb{R}^n \to \mathbb{R}^q$ that are assumed continuously differentiable. The notation $0 \leq u \perp v \geq 0$ for two vectors u and v in $\mathbb{R}^q$ is a shortcut for $u_i \geq 0,\ v_i \geq 0$ and $u_i v_i = 0$ for all $i \in \{1, \ldots, q\}$.

This problem has become an active subject in the literature in the last two decades and has been the subject of several monographs [56, 60] and PhD thesis [19, 45, 36, 69, 67, 15]. The wide variety of applications that can be cast as an MPCC is one of the reasons for this popularity. Among other we can cite truss topology optimization [36], discrete optimization [2], image restoration [12], optimal control [6, 35]. Otherwise, another source of problem are bilevel programming problems [17, 18], 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 [60] or Optimization Problem with Variational Inequality Constraint

*IRMAR-Insa, Rennes, France, e-mail: tangi.migot@insa-rennes.fr
†IRMAR-Insa, Rennes, France
‡BISOUS, Sherbrooke, Canada

[43]. 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 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 [43, 42, 65, 21] 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).

A wide range of numerical methods have been proposed to solve this problem such as relaxation methods, interior-point methods [55, 62, 52], penalty methods [56, 40, 58], SQP methods [26], elastic mode [10, 15], dc methods [57], filter methods [53] and Levenberg-Marquardt methods [33] to cite only a few of them. This first family of method called relaxation or regularization method is developped in detail in this paper. In the literature, the numerical performances of those methods are very often compared on a collection of test problem called MacMPEC, [51], that contains various (MPCC) from the literature and some applications. Despite the difficulties explained above, the methods reformulating (MPCC) as a non-linear programming have shown to be successful in practice in [24, 25]. However, the theoretical guarantee that they offer considering realistic methods is often far from the desired goal to compute M-stationary points.

In view of the constraint qualifications issues that pledge 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, as the relaxing parameter is reduced, convergence to the feasible set of (MPCC) is obtained similarly to an homotopy technique. The interest for such methods has already been the subject of some PhD thesis in [67, 69] and is an active subject in the literature.

These methods have been suggested in the literature back to 2000 by Scheel & Scholtes in [65] replacing the complementarity by

$$G_i(x)H_i(x) - t \leq 0, \ \forall i \in \{1, \dots, q\}.$$

For more clarity we denote $\Phi(G(x), H(x); t)$ the map that relaxed the complementarity constraint and so in this case

$$\Phi_i^{SS}(G(x), H(x); t) = G_i(x)H_i(x) - t, \ \forall i \in \{1, \dots, q\}. \tag{SS}$$

This natural approach was later extended by Demiguel, Friedlander, Nogales & Scholtes in [16] by also relaxing the positivity constraints $G(x) \geq -t$, $H(x) \geq -t$. In [54], 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 [68], the authors consider a

relaxation of the same type but only around the corner $G(x) = H(x) = 0$ in the following way

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

where ψ is a suitable function as described in [68]. An example of such function being $\psi(z) = \frac{2}{\pi}\sin(\frac{\pi}{2}z + \frac{3\pi}{2}) + 1$.

In the corresponding papers it has been shown that under suitable conditions providing convergence of the methods, convergence to some spurious point, called C-stationary point, may still happen. The convergence to M-stationary being guaranteed only under some second-order condition. 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 [46] providing an approximation scheme with convergence to M-stationary point by considering

$$\Phi_i^{KDB}(G(x), H(x); t) = (G_i(x) - t)(H_i(x) - t), \ \forall i \in \{1, \dots, q\}. \tag{KDB}$$

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 [48] :

$$\Phi_i^{KS}(G(x), H(x); t) = \phi(G_i(x) - t, H_i(x) - t), \ \forall i \in \{1, \dots, q\}. \tag{KS}$$

The main aim of this paper is to continue this discussion and extend the relaxation of Kanzow and Schwartz by introducing the new butterfly relaxation.

The key assumption necessary to guarantee convergence of the method relies very often on some MPCC-constraint qualification. In [49, 67] the authors analyse the existing methods and proves convergence under some mild constraint qualifications. The definition of a new MPCC-constraint qualification allows to pursue this discussion and convergence of (KDB) and (KS) has been shown under MPCC-CCP in [63]. Furthermore, the author proves that this is the weakest MPCC-constraint qualification that assures convergence of these methods. In this paper, we continue the discussion by providing convergence result for the butterfly method. MPCC-CCP condition is no longer sufficient for this purpose and so we introduce a new MPCC-constraint qualification called MPCC-CRSC.

In Section 2, we introduce classical definitions and results from non-linear programming and MPCC theory. This section is completed by the definition of a new constraint qualification for MPCC called MPCC-CRSC in Definition 2.11. This new concept is proved to be an MPCC-CQ in Corollary 2.2.

In Section 3, we define the relaxation scheme with the new butterfly relaxation. Compared to the methods (KDB) and (KS) this new method handle two relaxing parameter instead of one.

In Section 4, we prove theoretical results on convergence and existence of the multiplier of the relaxed sub-problems. We also provide an analysis on the convergence of approximate stationary points. We prove that the butterfly method has similar properties as the best methods in the literature.

Finally, in Section 5, we provide an extensive numerical study by giving detailed on the implementation, comparison with other methods as well as an example that illustrates the numerical difficulties that might occur and the extension of this method to the very close problem of mathematical program with vanishing constraints.

Notations : 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 $i \in \{1, \dots, n\}$. e is the vector whose components are all one. $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denotes the set of non-negative and positive real numbers.

2 Preliminaries

(MPCC) is obviously a non-linear programming problem. Most of numerical approaches used in non-linear programming compute necessary optimality conditions that require some constraint qualifications (CQs) defined in Sect. 2.1 to ensure existence of the Lagrange multipliers at a local minimum. Even so, (MPCC) belongs to this class of problem it is required to develop enhanced stationary conditions. Indeed, in a systematic way feasible points of (MPCC) may fail to satisfy even the weakest constraint qualifications for non-linear programming. Tailored optimality conditions and constraint qualifications for (MPCC) are presented in Sect. 2.2.

2.1 Non-Linear Programming

Let a general non-linear program be

$$\begin{array}{l} \min\limits_{x\in\mathbb{R}^n} f(x) \\ \text{s.t. } g(x)\leq 0,\ h(x)=0, \end{array} \tag{NLP}$$

with $h:\mathbb{R}^n\to\mathbb{R}^m$, $g:\mathbb{R}^n\to\mathbb{R}^p$ and $f:\mathbb{R}^n\to\mathbb{R}$.

Denote $\mathcal{F}$ the feasible region of (NLP), the set of active indices $\mathcal{I}_g(x):=\{i\in\{1,...,p\}\mid g_i(x)=0\}$. Let the generalized Lagrangian $\mathcal{L}^r(x,y)$ be

$$\mathcal{L}^r(x,y):=rf(x)+g(x)^T\lambda+h(x)^T\mu,$$

where $y=(\lambda,\mu)$ is the vector of Lagrange multiplier.

By definition, y is an index r multiplier for (NLP) at a feasible point x if $(r,y)\neq 0$ and $\nabla_x\mathcal{L}^r(x,y)=0,\ \lambda\geq 0,\ g(x)^T\lambda=0$. The set of all index r multipliers of (NLP) at x is denoted $\mathcal{M}^r(x)$. An index 0 multiplier is also called singular multiplier, [13], or an abnormal multiplier, [14]. We call a KKT-point or a stationary point a couple (x,y) with y an 1-index multiplier at x. A couple (x,y) with y a 0-index multiplier at x is called Fritz-John point.

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

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

Given a cone $K\subset\mathbb{R}^n$, the polar of K is the cone defined by $K^\circ:=\{z\in\mathbb{R}^n\mid z^Tx\leq 0,\ \forall x\in K\}$. Another useful tool for our study is the linearized cone of (NLP) at $x^*\in\mathcal{F}$ defined by

$$\mathscr{L}(x^*)=\{d\in\mathbb{R}^n\mid \nabla g_i(x)^Td\leq 0\ (i\in\mathcal{I}_g(x^*)),\ \nabla h_i(x)^Td=0\ (\forall i=1,\dots,m)\}.$$

In the context of solving non-linear programs, that is finding a local minimum of (NLP), one widely used technique is to compute necessary conditions. The principal tool is the Karush-Kuhn-Tucker (KKT) conditions. Let x^* be a local minimum of (NLP) that satisfies a constraint qualification, then there exists $y^*\in\mathcal{M}^1(x^*)$ such that (x^*,y^*) is a KKT-point of (NLP). Constraint qualification are used to ensure existence of the index-1 multiplier at x^*.

We now define some of the classical constraint qualifications that are organized in several families. Note that there exists a wide variety of such notions and we define here those that are essential for our purpose. In Definition 2.1 the Linear Independence CQ and Constant Rank CQ are presented. Both are very classical the latter being defined first in [44].

Definition 2.1 (LICQ and its Relaxations)**.** *Let $x^*\in\mathcal{F}$.*

(a) LICQ holds at x^* if the family of gradients

$$\{\nabla g_i(x^*)\ (i\in\mathcal{I}_g(x^*)),\ \nabla h_i(x^*)\ (\forall i=1,...,m)\}$$

is linearly independent.

(b) CRCQ holds at x^* if there exists $\delta > 0$ such that, for any subsets $I_1 \subseteq \mathcal{I}_g(x^*)$ and $I_2 \subseteq \{1, ..., m\}$ the family of gradients

$$\{\nabla g_i(x) \ (i \in I_1), \ \nabla h_i(x) \ (i \in I_2)\}$$

has the same rank for all $x \in \mathcal{B}_\delta(x^*)$.

We remind the following definition of positive-linearly dependent vectors, which helps up building constraint qualifications since it takes into account the sign of some multipliers.

Definition 2.2 (positive-linearly dependent vectors)**.** *A finite set of vectors $\{a^i | i \in I_1\} \cup \{b^i | i \in I_2\}$ is said to be positive-linearly dependent if there exist scalars α_i $(i \in I_1)$ and β_i $(i \in I_2)$, not all of them being zero, with $\alpha_i \geq 0$ for all $i \in I_1$ and*

$$\sum_{i \in I_1} \alpha_i a^i + \sum_{i \in I_2} \beta_i b^i = 0.$$

Otherwise, we say that these vectors are positive-linearly independent. In an usual way the unsigned vectors may sometimes be denoted with double bracket, that is $\{a^i | i \in I_1\} \cup \{\{b^i | i \in I_2\}\}$.

Another family of constraint qualifications can now be derived using this notion.

Definition 2.3 (MFCQ and its Relaxations)**.** *Let $x^* \in \mathcal{F}$.*

(a) PLICQ holds at x^* if the gradients

$$\{\nabla g_i(x^*) \mid i \in \mathcal{I}_g(x^*)\} \cup \{\{\nabla h_j(x^*) \mid j = 1, ..., m\}\}$$

are positively linearly independent.

(b) CRSC holds at x^* if there exists $\delta > 0$ such that the family of gradients

$$\{\nabla g_i(x) \ (i \in J_-), \ \nabla h_i(x^*) \ (i = 1, \ldots, m)\}$$

has the same rank for every $x \in \mathcal{B}_\delta(x^*)$. Assuming $J_- := \{i \in \mathcal{I}_g(x^*) \mid -\nabla g_i(x^*) \in \mathscr{L}(x^*)^\circ\}$.

The positive linear independence constraint qualification, PLICQ, is equivalent to MFCQ and has also been called NNAMCQ (no nonzero abnormal multiplier) or BCQ (basic constraint qualification). Constant rank of the subspace component, CRSC, was introduced recently in [8]. This latter definition consider an unusual set denoted J_-, that can be view as the set of indices of the gradients of the active constraints whose Lagrange multiplier if there exists may be nonzero.

A local minimum is characterized by the fact that there is no feasible descent direction for the objective function of (NLP), that is

$$-\nabla f(x^*) \in \mathcal{T}_\mathcal{F}^\circ(x^*).$$

From the other side the KKT conditions build ∇f using a linearization of the active constraints. This motivates the following CQs defined as early as 1969 in [31] for GCQ and in [1] for ACQ.

Definition 2.4. *(a) A point $x^* \in \mathcal{F}$ is said to satisfy Guignard CQ if $\mathcal{T}_\mathcal{F}^\circ(x^*) = \mathscr{L}^\circ(x^*)$.*

(b) A point $x^* \in \mathcal{F}$ is said to satisfy Abadie CQ if $\mathcal{T}_\mathcal{F}(x^*) = \mathscr{L}(x^*)$.

As proved in [30], Guignard CQ is the weakest constraint qualification that ensures that a local minimum satisfies the KKT conditions. It is easy to see that for all x $\mathcal{T}_\mathcal{F}(x) \subseteq \mathscr{L}(x)$ and so $\mathscr{L}^\circ(x) \subseteq \mathcal{T}_\mathcal{F}^\circ(x)$. The fact that Abadie CQ holds at x^* implies that Guignard CQ also holds at x^* is classical from variational analysis.

In practice, it is very difficult to find a point that conforms exactly to the KKT condition. Hence, an algorithm may stop when such conditions are satisfied approximately. Another way to deal with this problem is to gives necessary optimality conditions of the point and its neighbourhood in the form of sequential optimality conditions. The most popular among those conditions is the Approximate KKT (AKKT) conditions introduced in [7]. This has motivated the definition of the CCP condition in [9] that is known to be the weakest constraint qualification that ensures that AKKT is actually a first order optimality condition.

Definition 2.5. *We say that a point $x^* \in \mathcal{F}$ satisfies the Cone-Continuity Property if the set-valued mapping $\mathbb{R}^n \ni x \rightrightarrows K(x)$ such that*

$$K(x) = \{ \sum_{i \in \mathcal{I}_g(x^*)} \lambda_i \nabla g_i(x) + \sum_{i=1}^{m} \mu_i \nabla h_i(x) \ : \ \lambda_i \in \mathbb{R}_+, \mu_i \in \mathbb{R} \}$$

is outer semicontinuous (definition 5.4 [64]) at x^*, that is

$$\limsup_{x \to x^*} K(x) \subset K(x^*).$$

Clearly, $K(x^*)$ is a closed convex cone and coincides with the polar linearized cone $\mathscr{L}(x^*)^\circ$. Moreover, $K(x)$ is always inner semicontinuous due to the continuity of the gradients and the definition of $K(x)$. For this reason, outer semicontinuity is sufficient for the continuity of $K(x)$ at x^*. Finally it has been shown in [9] that CCP is stricly stronger than ACQ and weaker than CRSC.

In the context of numerical computation it is almost never possible to compute stationary points. Hence, it is of interest to consider ϵ-stationary points.

Definition 2.6. *Given a general non-linear program* (NLP) *and $\epsilon \geq 0$. We say that $(x, y) \in \mathbb{R}^n \times \mathbb{R}^{p+m}$ is an ϵ-stationary point (or an ϵ-KKT point) if it satisfies*

$$\left\| \nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) + \sum_{i=1}^{m} \mu_i \nabla h_i(x) \right\|_\infty \leq \epsilon,$$

with

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

2.2 Mathematical Program with Complementarity Constraints

We now specialize the general notions above to our specific case of (MPCC). Let $\mathcal{Z}$ be the set of feasible points of (MPCC). Given $x^* \in \mathcal{Z}$, we denote

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

As in the previous section, we extend the definition of the Lagrangian. Let $\mathcal{L}^r_{MPCC}$ be the generalized MPCC-Lagrangian function of (MPCC) such that

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

with $\lambda := (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$.

2.2.1 Stationary Point

It is clear that we can not expect to compute usual KKT-point since classical constraint qualifications in general does not hold, so we introduce weaker stationary concept as in [65, 42].

Definition 2.7. *$x^* \in \mathcal{Z}$ is said*

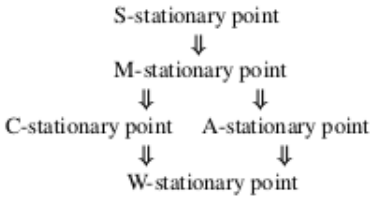

Figure 1: Relations between the stationarity definitions

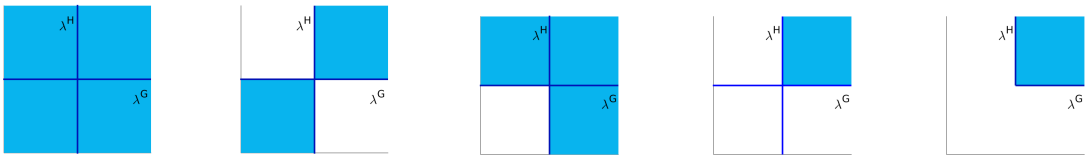

Figure 2: Signs of λ^G, λ^H for indices $i \in \mathcal{I}^{00}$. From the left to the right : weak-stationary, C-stationary, A-stationary, M-stationary and S-stationay.

- *Weak-stationary if there exists* $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in \mathbb{R}^p_+ \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$ *such that*

$$\nabla_x \mathcal{L}^1_{MPCC}(x^*, \lambda^g, \lambda^h, \lambda^G, \lambda^H) = 0,$$
$$\lambda^g_i = 0 \ \forall i \notin \mathcal{I}_g, \ \lambda^G_{\mathcal{I}^{+0}(x^*)} = 0, \ \lambda^H_{\mathcal{I}^{0+}(x^*)} = 0.$$

- *Clarke (C)-stationary point if* x^* *is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \ \lambda^G_i \lambda^H_i \geq 0.$$

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

$$\forall i \in \mathcal{I}^{00}(x^*), \ \lambda^G_i \geq 0 \text{ or } \lambda^H_i \geq 0.$$

- *Mordukhovich (M)-stationary point if* x^* *is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \text{ either } \lambda^G_i > 0, \ \lambda^H_i > 0 \text{ or } \lambda^G_i \lambda^H_i = 0.$$

- *Strong (S)-stationary point if* x^* *is weak-stationary and*

$$\forall i \in \mathcal{I}^{00}(x^*), \ \lambda^G_i \geq 0, \ \lambda^H_i \geq 0.$$

Relations between these definitions are given in Figure 1 and follow in a straightforward way from the definitions. Local optimal solution are often denoted Bouligand (B)-stationary point in the literature, but this will not be used here.

2.2.2 First Order Constraint Qualification for MPCC

In a classical way from the literature, we extend the various constraint qualifications for (NLP) to (MPCC). MPCC-CQ denotes this extension of usual CQ.

Abadie CQ and Guignard CQ are the weakest constraint qualifications in non-linear programming. Unfortunately Abadie condition is very unlikely to be satisfied with (MPCC). Indeed, the tangent cone, $\mathcal{T}_{\mathcal{Z}}$, is

closed but in general not convex and the classical linearized cone of (MPCC) is polyhedral for (MPCC) and therefore convex. That is why we define a specific cone for (MPCC) denoted $\mathscr{L}_{MPCC}$ as in [65, 20, 61]

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

This cone is no longer a polyhedral cone and is not necessarily convex. However due to [20], one always has the following inclusions

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

Definition 2.8. *Let $x^* \in \mathcal{Z}$. We say that MPCC-ACQ holds at x^* if $\mathcal{T}_{\mathcal{Z}}(x^*) = \mathscr{L}_{MPCC}(x^*)$ and MPCC-GCQ holds at x^* if $\mathcal{T}^\circ_{\mathcal{Z}}(x^*) = \mathscr{L}^\circ_{MPCC}(x^*)$.*

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

Theorem 2.1 ([23])**.** *A local minimum of (MPCC) that satisfies MPCC-GCQ or any stronger MPCC-CQ is an M-stationary point.*

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

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

As a consequence of the previous theorem we can deduce the following result.

Lemma 2.1. *Let $x^* \in \mathcal{Z}$ such that MPCC-GCQ holds at x^*. Then, the following inclusion holds true*

$$\mathscr{L}^\circ_{MPCC}(x^*) \subset \mathscr{P}_M.$$

Proof. First, it is easy to notice the following characterization of M-stationary point with the $\mathscr{P}_M$. x^* is an M-stationary point if and only if

$$-\nabla f(x^*) \in \mathscr{P}_M. \tag{2}$$

Theorem 2.1 proved that given $x^* \in \mathcal{Z}$ such that MPCC-GCQ holds true at x^*, then for any continuously differentiable function f it follows

$$\begin{aligned}x^* \text{ is a local minimum } \Longleftrightarrow -\nabla f(x^*) \in \mathcal{T}^\circ_{\mathcal{Z}}(x^*) \quad &\Longleftrightarrow \quad -\nabla f(x^*) \in \mathscr{L}^\circ_{MPCC}(x^*)\\ &\Longrightarrow \quad x^* \text{ is an M-stationary point}\\ &\Longleftrightarrow \quad -\nabla f(x^*) \in \mathscr{P}_M.\end{aligned}$$

The last equation comes from the characterization of M-stationary point given in (2).

Since f is taken in a completely general way, this proves the result. □

One of the most principal constraint qualification used in the literature of (MPCC) is the MPCC-LICQ, see [66] for a discussion on this CQ. In a similar way we extend CRCQ as in [32]. A condition that is similar was used in [48, 39] to prove convergence of relaxation methods for (MPCC).

Definition 2.9. *Let $x^* \in \mathcal{Z}$.*

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

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

   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 $x \in \mathcal{B}_\delta(x^*)$.

The linear independance CQ in the context has a very specific behaviour since in this case KKT conditions holds as stated in the following result from [22] (Theorem 4.5).

Theorem 2.2. *If a point $x^* \in \mathcal{Z}$ satisfies MPCC-LICQ it also satisfies classical GCQ.*

In [22], they also provide examples to show than GCQ does not hold with weaker MPCC-CQs. So, under MPCC-LICQ the correct stationary concept is S-stationary and therefore M-stationary points may be undesirables.

However, as pointed out in Theorem 2.1, in the general case the correct sign of the multiplier λ_i^G, λ_i^H for $i \in \mathcal{I}^{00}(x^*)$ in the necessary optimality conditions for (MPCC) are the sign of M-stationary points. This motivates the following definition of MPCC-MFCQ that specialize the MPCC-LICQ by taking into account those signs of multipliers for $i \in \mathcal{I}^{00}(x^*)$.

Definition 2.10. *Let $x^* \in \mathcal{Z}$. MPCC-MFCQ holds at x^* if the only solution of*

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

with $\lambda_i^g \geq 0$ $(i \in \mathcal{I}_g(x^*))$ and either $\lambda_i^G \lambda_i^H = 0$ either $\lambda_i^G > 0, \lambda_i^H > 0$ for all $i \in \mathcal{I}^{00}(x^*)$ is the trivial solution.

2.3 A New MPCC-Constraint Qualification

In a similar way as for MPCC-MFCQ, we extend the definition of CRSC constraint qualification to introduce the MPCC-CRSC, which is new in the MPCC literature.

Definition 2.11. *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*

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

with $\lambda_i^g \geq 0$ $(i \in \mathcal{I}_g(x^*))$,$\lambda_i^G$ 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

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

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

In the special case where there is no partition of $\mathcal{I}^{00}(x^*)$ that satisfies the condition of the definition above, then obviously the gradients are linearly independant and so MPCC-MFCQ holds at x^*.

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.

We now state that this new notion of MPCC-CRSC is actually an MPCC-CQ by proving that it implies MPCC-CCP.

Definition 2.12. *We say that a feasible point x^* satisfies the MPCC-CCP if the set-valued mapping $\mathbb{R}^n \ni x \rightrightarrows K_{MPCC}(x)$ such that*

$$K_{MPCC}(x) := \{ \sum_{i \in \mathcal{I}_g(x^*)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^m \lambda^h_i \nabla h_i(x) - \sum_{i \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda^G_i \nabla G_i(x) - \sum_{i \in \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda^H_i \nabla H_i(x)$$
$$: \lambda^g_i \in \mathbb{R}_+ \text{ and , either } \lambda^G_i \lambda^H_i = 0 \text{ either } \lambda^G_i > 0, \lambda^H_i > 0 \text{ for } i \in \mathcal{I}^{00}(x^*)\}$$

is outer semicontinuous (Definition 5.4 [64]) at x^*, that is

$$\limsup_{x \to x^*} K_{MPCC}(x) \subset K_{MPCC}(x^*).$$

This definition is motivated by sequential optimality conditions from [9] for non-linear programming and extended for (MPCC) in [63], where it has been proved to be an MPCC-constraint qualification.

A useful lemma that we extensively use in the following results is reminded here. This result is a Caratheodory kind lemma.

Lemma 2.2. *[Lemma 7.1, [67]] Let $\{a_i \mid i = 1, \dots, p\}$, $\{b_i \mid i = 1, \dots, m\}$ and c be vectors in $\mathbb{R}^n$ and $\alpha \in \mathbb{R}^p_+$, $\beta \in \mathbb{R}^m$ multipliers such that*

$$\sum_{i=1}^p \alpha_i a^i + \sum_{i=1}^m \beta_i b^i = c.$$

Then there exist multipliers $\alpha^* \in \mathbb{R}^p_+$ and β^* with $supp(\alpha^*) \subseteq supp(\alpha)$, $supp(\beta^*) \subseteq supp(\beta)$ and

$$\sum_{i=1}^p \alpha^*_i a^i + \sum_{i=1}^m \beta^*_i b^i = c$$

such that the vectors

$$\{a^i \mid i \in supp(\alpha^*)\} \cup \{b^i \mid i \in supp(\beta^*)\}$$

are linearly independent.

The following results give a characterization of some sequences that satisfy MPCC-CRCQ and MPCC-CRSC at their limit point. Note that this result is essential for the convergence proof of relaxation methods for (MPCC) that will be studied in the next section, since it proves boundedness of approximate stationary sequences.

Theorem 2.3. *Let x^* be in $\mathcal{Z}$ such that MPCC-CRCQ holds at x^*. Given two sequences $\{x^k\}$,$\{\lambda^k\}$ so that up to a subsequence $x^k \to x^*$ and λ^k goes to some limit $\lambda \in \mathbb{R}^p_+ \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$ (possibly infinite) that satisfies*

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) - \sum_{i=1}^{q} \lambda_i^{G,k} \nabla G_i(x^k) - \sum_{i=1}^{q} \lambda_i^{H,k} \nabla H_i(x^k) \to 0, \tag{3}$$

and

$$\lambda_i^g = 0 \ \forall i \notin \mathcal{I}_g(x^*), \ \forall i \in \mathcal{I}^{+0}(x^*) \ \lim_{k\to\infty} \frac{\lambda_i^{G,k}}{\|\lambda^k\|_\infty} = 0 \text{ and } \forall i \in \mathcal{I}^{0+}(x^*) \ \lim_{k\to\infty} \frac{\lambda_i^{H,k}}{\|\lambda^k\|_\infty} = 0.$$

Then the sequence $\{\lambda^k\}$ is bounded.

Proof. Let $\{w^k\}$ be a sequence defined such that

$$w^k := \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) + \sum_{j \in \mathcal{I}_g(x^*)} \lambda_j^{g,k} \nabla g_j(x^k) - \sum_{j \in \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda_j^{G,k} \nabla G_j(x^k) - \sum_{j \in \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*)} \lambda_j^{H,k} \nabla H_j(x^k).$$

We can safely assume that for k sufficiently large $w^k \neq 0$. Indeed, if up to some indice $\bar{k}$ it holds that $w^k = 0$ for all $k \geq \bar{k}$ then assumption (3) implies that $\nabla f(x^k) \to \nabla f(x^*) = 0$. So, in this case the constraints of the problem are not playing any role and for $k \geq \bar{k}$, $\lambda^k = 0$ is a valid sequence of Lagrange multiplier.

According to Lemma 2.2 we may assume without loss of generality that the gradients corresponding to non-vanishing multipliers in the definition of w^k are linearly independent for all $k \in \mathbb{N}$. (note that this may change the multipliers, but a previously positive multiplier will stay at least non-negative and a vanishing multiplier will remain zero).

We prove by contradiction that the sequence $\{\lambda^k\}$ is bounded. Assuming that for some indices λ^k is not bounded, therefore there exists a subsequence such that

$$\frac{\lambda^k}{\|\lambda^k\|_\infty} \to \lambda \neq 0.$$

Dividing by $\|\lambda^k\|_\infty$ and passing to the limit in the equation above yields

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

with $\lambda_j^G = 0$ for $j \in \mathcal{I}^{+0}$, $\lambda_j^H = 0$ for $j \in \mathcal{I}^{0+}$ by assumption (3).

It follows that the gradients with non-zero multipliers involved in the previous equation are linearly dependent. MPCC-CRCQ guarantees that these gradients remain linearly dependent in a whole neighbourhood. This, however, is a contradiction to the linear independence of these gradients in x^k since $w^k \neq 0$ for k sufficiently large. Here, we used that for all k sufficiently large $\text{supp}(\lambda) \subseteq \text{supp}(\lambda^k)$.

Consequently, our assumption was wrong and thus the sequence $\{\lambda^k\}$ is bounded. □

The following result is a consequence of Theorem 2.3 in the case where the limit point is an M-stationary point.

Corollary 2.1. *Let x^* be in $\mathcal{Z}$ such that MPCC-CRSC holds at x^*. Given two sequences $\{x^k\}$,$\{\lambda^k\}$ so that up to a subsequence $x^k \to x^*$ and λ^k goes to some limit $\lambda \in \mathbb{R}^p_+ \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$ (possibly infinite) that satisfies*

$$\nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) - \sum_{i=1}^{q} \lambda_i^{G,k} \nabla G_i(x^k) - \sum_{i=1}^{q} \lambda_i^{H,k} \nabla H_i(x^k) \to 0, \tag{4}$$

$$\lambda_i^g = 0 \ \forall i \notin \mathcal{I}_g(x^*),\ \forall i \in \mathcal{I}^{+0}(x^*)\ \lim_{k\to\infty} \frac{\lambda_i^{G,k}}{\|\lambda^k\|_\infty} = 0 \text{ and } \forall i \in \mathcal{I}^{0+}(x^*)\ \lim_{k\to\infty} \frac{\lambda_i^{H,k}}{\|\lambda^k\|_\infty} = 0,$$

and

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

Then the sequence $\{\lambda^k\}$ is bounded.

Proof. The proof is completely similar to Theorem 2.3. It remains to verify that under the assumption MPCC-CRSC the gradients of non-vanishing multipliers involved in the following equation

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

are linearly dependent in whole neighbourhood of x^*.

However, it is clear that the family of gradients considered in the definition of MPCC-CRSC corresponds to the gradients with non-zero multipliers in the previous equation. Since by linear dependence of the gradients at x^* any gradient whose multiplier is non-zero may be formulated as a linear combination of the others gradients. Therefore, those gradients with non-vanishing multipliers belong to the polar of the M-linearized cone. MPCC-CRSC guarantees that these gradients remain linearly dependent in a whole neighbourhood.

Applying Theorem 2.3 the sequence $\{\lambda^k\}$ is bounded. □

We conclude this section by a consequence of Corollary 2.1 that states an essential result for this section, namely MPCC-CRSC is an MPCC-constraint qualification.

Corollary 2.2. *MPCC-CRSC implies MPCC-CCP.*

Proof. We prove that a point x^* that satisfies MPCC-CRSC satisfies the following relation

$$\limsup_{x\to x^*} K_{MPCC}(x) \subset K_{MPCC}(x^*).$$

Let w^* be in the $\limsup_{x\to x^*} K_{MPCC}(x)$. By definition of the lim sup there are sequences $\{w^k\}, \{x^k\}, \{\lambda^k\}$ with $x^k \to x^*$ and $w^k \to w^*$ such that for k sufficiently large

$$w^k = \sum_{i=1}^m \lambda_i^{h,k} \nabla h_i(x^k) + \sum_{j\in\mathcal{I}_g(x^*)} \lambda_j^{g,k} \nabla g_j(x^k) - \sum_{j\in\mathcal{I}^{0+}(x^*)\cup\mathcal{I}^{00}(x^*)} \lambda_j^{G,k} \nabla G_j(x^k) - \sum_{j\in\mathcal{I}^{+0}(x^*)\cup\mathcal{I}^{00}(x^*)} \lambda_j^{H,k} \nabla H_j(x^k),$$

with $\lambda_i^{g,k} \in \mathbb{R}_+$, either $\lambda_i^{G,k}\lambda_i^{H,k} = 0$ either $\lambda_i^{G,k} > 0, \lambda_i^{H,k} > 0$ for $i \in \mathcal{I}^{00}(x^*)$. Moreover, for k sufficiently large it holds that $\text{supp}(\lambda^{g,k}) \subseteq \mathcal{I}_g(x^*)$, $\text{supp}(\lambda^{G,k}) \subseteq \mathcal{I}^{0+}(x^*) \cup \mathcal{I}^{00}(x^*)$ and $\text{supp}(\lambda^{H,k}) \subseteq \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{+0}(x^*)$.

The sequence $\{\lambda^k\}$ clearly satisfies the assumption (4) of Corollary 2.1. It follows that this sequence is bounded and up to a subsequence we can extract a limit point λ^*. Consequently, by definition of λ^k it holds that $\lambda_j^{G,*} = 0$ for $j \in \mathcal{I}^{+0}$, $\lambda_j^{H,*} = 0$ for $j \in \mathcal{I}^{0+}$ and either $\lambda_j^{G,*}\lambda_j^{H,*} = 0$ either $\lambda_j^{G,*} > 0, \lambda_j^{H,*} > 0$ for $j \in \mathcal{I}^{00}(x^*)$.

So, we can conclude that w^* belongs to $K_{MPCC}(x^*)$ and therefore MPCC-CCP is satisfied at x^*. □

An alternative proof could used Theorem 3.5 of [63] that state that any sequence $\{x^k\}$ that converges to an M-stationary point x^* is equivalent to satisfying MPCC-CCP at x^*. Note that this is conceptually equivalent to the one presented here.

We sum up this section in Figure 3 by giving the relationship between the various MPCC-CQ defined here.

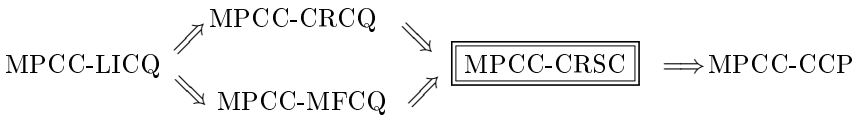

Figure 3: Relation between the MPCC-constraint qualifications

3 The Butterfly Relaxation Methods

3.1 Relaxation Methods

The principal aim of this paper focuses on relaxation methods to solve (MPCC). The sketch of such a method is described in Algorithm 1 and behaves as follows: we consider a non-linear parametric program R_{t_k}, where the complementarity constraints have been relaxed using a parameter $t_k > 0$. A sequence $\{x^{k+1}\}$ of stationary points of R_{t_k} is then computed for each value of $t_k > 0$. Such stationary points are computed using iterative methods that require an initial point. We use the previous stationary point as an initial point. For t_k converging to zero the sequence $\{x^{k+1}\}$ converge under some mild assumption to a stationary point of (MPCC).

The relaxed problems R_{t_k} with a parameter t_k are written in a general form as

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\quad g(x) \leq 0,\ h(x) = 0, \\ &\quad G(x) \geq 0,\ H(x) \geq 0, \\ &\quad \Phi_i(G(x), H(x); t_k) \leq 0,\ i = 1, \ldots, q\ . \end{aligned} \tag{R_t}$$

We may skip the indices k in the notation of the parameter t_k, when it is clear from the context that we look at the step k.

Data:
starting vector x^0; sequence of relaxation parameters $\{t_k\}$;
Begin ;
Set $k := 0$;
while $x^k \notin \mathcal{Z}$ **do**
  x^{k+1}:= stationary point of $R(t_k)$ with x^k as an initial point;

Algorithm 1: Generic relaxation method for (MPCC), with a corresponding relaxed non-linear program R_t.

According to Section 2.2, our aim is to compute an M-stationary point of (MPCC). A motivation to consider such method is that the sequence of relaxed non-linear program may satisfy some constraint qualification and then are more tractable for classical non-linear methods.

The following section introduces our new relaxation schemes called butterfly relaxations.

3.2 Butterfy Relaxations

We propose a new family of relaxations with two positive parameters (t, r) defined such that for all $i \in \{1, \ldots, q\}$

$$\Phi_i^B(G(x), H(x); t, r) := \begin{cases} F_{1i}(x; r, t)F_{2i}(x; t, r), & \text{if } F_{1i}(x; r, t)F_{2i}(x; t, r) \geq 0 \\ < 0, & \text{otherwise} \end{cases} \tag{5}$$

with

$$\begin{aligned} F_{1i}(x; r, t) &:= (H_i(x) - t\theta_r(G_i(x))) \\ F_{2i}(x; t, r) &:= (G_i(x) - t\theta_r(H_i(x)))' \end{aligned}$$

where $\theta_r : \mathbb{R} \to]-\infty, 1]$ are continuously differentiable non-decreasing concave functions with $\theta(0) = 0$, and $\lim_{r\to 0} \theta_r(x) = 1 \ \forall x \in \mathbb{R}_{++}$ completed in a smooth way for negative values by considering $\theta_r(z < 0) = \frac{z\theta'(0)}{r}$. Examples of such functions are

$$\theta_r^1(x) = \frac{x}{x+r} \text{ or } \theta_r^2(x) = 1 - \exp^{-\frac{x}{r}}.$$

Those functions have already been used in the context of complementarity constraints in [34]. These new relaxations handle two parameters r and t instead of one and chosen such that

$$t\theta'(0) \leq r \text{ and } t = \omega(r^2). \tag{6}$$

One way to write (5) for $t < \theta'(0)r$ uses the NCP function from [67] by considering

$$\Phi_i^B(G(x), H(x); t, r) := \begin{cases} F_{1i}(x; r, t)F_{2i}(x; t, r), & \text{if } F_{1i}(x; r, t) + F_{2i}(x; t, r) \geq 0 \\ -\frac{1}{2}(F_{1i}(x; r, t)^2 + F_{2i}(x; t, r)^2), & \text{if } F_{1i}(x; r, t) + F_{2i}(x; t, r) < 0. \end{cases} \tag{7}$$

Since these relaxations are an union of two convex sets connected on a single point we may also consider a relaxation of the positivity constraints. The feasible set of the relaxed complementarity constraints is presented in Figure 4 for some examples. This method is an extension of the work of [46, 47] and [38, 48] since handling two parameters allows bringing the two "wings" of the relaxation closer.

We now introduce some notations that will be extensively used in the sequel. Since the butterfly relaxations handle two parameters we denote $\hat{t} := (t, r)$ to simplify the notation and by extension $\hat{t}_k := (t_k, r_k)$. Let $\mathcal{X}_{\hat{t}}^B$ be the feasible set of $R_{\hat{t}}^B$, which corresponds to the non-linear program related to the butterfly relaxation of the complementarity constraints defined in (5), that is

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{s.t } g(x) \leq 0,\ h(x) = 0, \\ &\quad G(x) \geq -\bar{r}e,\ H(x) \geq -\bar{r}e, \\ &\quad \Phi_i^B(G(x), H(x); \hat{t})(x) \leq 0,\ i = 1, \ldots, q\ . \end{aligned} \tag{$R_{\hat{t}}^B$}$$

The parameter $\bar{r}$ depends on t and r and is chosen such that

$$\{x \in \mathbb{R}^n \mid -\bar{r}e \leq G(x), H(x) \leq 0\} \subset \{x \in \mathbb{R}^n \mid F_{1i}(x; r, t) + F_{2i}(x; t, r) \leq 0,\ i = 1, \ldots, q\}.$$

As an example, one can choose $\bar{r} := -\frac{2r^2(1-\theta'(0)t/r)}{t\theta''(0)}$. It is to be noted that $\bar{r}$ is positive by (6). Later, we discuss the asymptotic behaviour of this parameter and it is of interest to note that when $t = o(r)$ then $\bar{r}$ goes to 0 in the same order than r^2/t, we denote $\bar{r} \sim_K r^2/t$. Figure 4 shows the feasible set of the relaxed complementarity constraint for some relations between t and r.

The sets of indices used in the sequel are defined in the following way

$$\begin{aligned} \mathcal{I}_G(x; \hat{t}) &:= \{i = 1, \ldots, q \mid G_i(x) + \bar{r} = 0\} \\ \mathcal{I}_H(x; \hat{t}) &:= \{i = 1, \ldots, q \mid H_i(x) + \bar{r} = 0\} \\ \mathcal{I}_{GH}(x; \hat{t}) &:= \{i = 1, \ldots, q \mid \Phi_i^B(G(x), H(x); \hat{t})(x) = 0\} \\ \mathcal{I}_{GH}^{0+}(x; \hat{t}) &:= \{i \in \mathcal{I}_{GH}(x; \hat{t}) \mid F_{1i}(x; \hat{t}) = 0,\ F_{2i}(x; \hat{t}) > 0\} \\ \mathcal{I}_{GH}^{+0}(x; \hat{t}) &:= \{i \in \mathcal{I}_{GH}(x; \hat{t}) \mid F_{1i}(x; \hat{t}) > 0,\ F_{2i}(x; \hat{t}) = 0\} \\ \mathcal{I}_{GH}^{++}(x; \hat{t}) &:= \{i \in \mathcal{I}_{GH}(x; \hat{t}) \mid F_{1i}(x; \hat{t}) > 0,\ F_{2i}(x; \hat{t}) > 0\} \\ \mathcal{I}_{GH}^{00}(x; \hat{t}) &:= \{i \in \mathcal{I}_{GH}(x; \hat{t}) \mid F_{1i}(x; \hat{t}) = F_{2i}(x; \hat{t}) = 0\}. \end{aligned}$$

We also use classical asymptotic Landau notations :

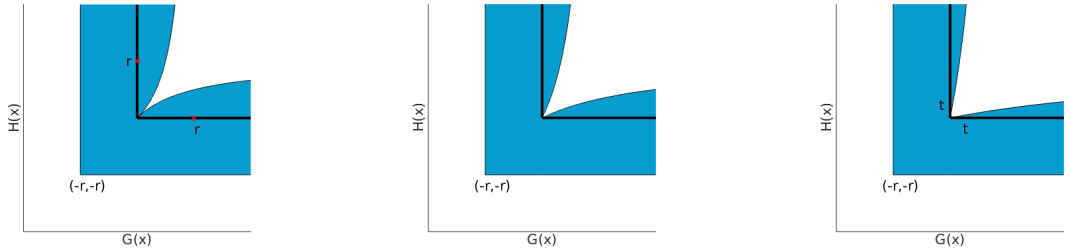

Figure 4: Feasible set of the butterfly relaxation for $\theta_r(z) = \frac{z}{z+r}$ with from the left to the right : $t = r$, $t = 2r$ and $t = r^{3/2}$.

$f(x) = O(g(x))$ as $x \to a$ if and only if there exists positive numbers δ and M such that $|f(x)| \leq M|g(x)|$ for all $|x - a| \leq \delta$.

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

$f(x) = o(g(x))$ as $x \to a$ 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 $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$.

$f(x) = \omega(g(x))$ as $x \to a$ 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 $|x - a| \leq \delta$.

$f(x) \sim_K (g(x))$ as $x \to a$ if and only $\lim_{x\to a} \frac{f(x)}{g(x)} = K$ with K a positive finite constant and in a classical way $f(x) \sim_K (g(x))$ when $K = 1$.

We focus in the sequel on the properties of these new relaxation schemes. The asymptotic behaviour of these methods are illustrated in Lemmas 3.2 and 3.3. Then, we move to convergence properties of the methods considering a sequence of stationary points, which is proved to converge to an A-stationary point in Theorem 4.1 and to an M-stationary point, Corollary 4.1, with some relation between the sequences $\{t_k\}$ and $\{r_k\}$. In both results we discuss weak constraint qualifications necessary to ensure convergence of these methods. The main motivation to consider relaxation methods for (MPCC) is to solve a sequence of regular problems. Under classical assumptions the butterfly relaxed non-linear programs satisfy Guignard CQ, Theorem 4.2. A specific kind of butterfly methods, where $t_k\theta'(0) = r_k$ for k sufficiently large, has improved properties since they satisfy Abadie CQ, Theorem 4.3. It is more realistic to consider a sequence of ϵ-stationary points instead of classical stationary points. Also there is a price to pay here since convergence properties of most of the relaxation methods are damaged. This is discussed in Theorem 4.5 for the formulation with $t_k = o(r_k)$. Finally, numerical results will be presented in Sect. 5 and show that these new methods are very competitive compared to existing methods.

Before moving to our main results regarding convergence and regularity properties of the butterfly relaxations, we provide some useful results on the asymptotic behaviour of functions θ_r and $\Phi^B(G(x), H(x); \hat{t})$. Direct computation gives the gradient of $\Phi^B(G(x), H(x); \hat{t})$ in the following lemma.

Lemma 3.1. *For all $i \in \{1, \ldots, q\}$, the gradient of $\Phi_i^B(G(x), H(x); \hat{t})$ at x is given by*

$$\nabla\Phi_i^B(G(x), H(x); \hat{t})(x) = \begin{cases} \left(F_{1i}(x; \hat{t}) - t\theta_r'(G_i(x))F_{2i}(x; \hat{t})\right) \nabla G_i(x) \\ + \left(F_{2i}(x; \hat{t}) - t\theta_r'(H_i(x))F_{1i}(x; \hat{t})\right) \nabla H_i(x) \text{ if } F_{1i}(x; \hat{t}) + F_{2i}(x; \hat{t}) \geq 0 \\ \left(t\theta_r'(G_i(x))F_{1i}(x; \hat{t}) - F_{2i}(x; \hat{t})\right) \nabla G_i(x) \\ + \left(t\theta_r'(H_i(x))F_{2i}(x; \hat{t}) - F_{1i}(x; \hat{t})\right) \nabla H_i(x) \text{ if } F_{1i}(x; \hat{t}) + F_{2i}(x; \hat{t}) < 0 \end{cases}$$

The following result illustrates the behaviour of functions θ_r and their derivatives when t and r are going through zero.

Lemma 3.2. *Given two sequences* $\{r_k\}$ *and* $\{t_k\}$*, which converge to* 0 *as* k *goes to infinity and* $\forall k \in \mathbb{N},\ (r_k, t_k) \in \mathbb{R}^2_{++}$*. Then, for any* $z \in \mathbb{R}_+$

$$\lim_{k\to\infty} t_k \theta_{r_k}(z) = 0.$$

Furthermore, let $\{z^k\}$ *be such that* $\lim_{k\to\infty} z^k = 0$*. Then, either* $z^k = O(r_k)$ *and so there exists a constant* $C_\theta \in [0, \theta'(0)]$ *such that,*

$$\lim_{k\to\infty} t_k \theta'_{r_k}(z_k) = \lim_{k\to\infty} C_\theta \frac{t_k}{r_k},$$

otherwise, i.e $z^k = \omega(r_k)$*, then*

$$\lim_{k\to\infty} t_k \theta'_{r_k}(z_k) \leq \lim_{k\to\infty} \theta'(1) \frac{t_k}{r_k}.$$

Proof. First part of the lemma follows from the definition of functions θ_r. Indeed, it holds for all $z \in \mathbb{R}_+$ that $\theta_r(z) \in [0, 1]$. Therefore, $\lim_{k\to\infty} t_k \theta_{r_k}(z_k) = 0$.

Second part of the lemma uses the fact that functions θ_r are defined as perspective functions, that is for all $z^k \in \mathbb{R}_+$

$$\theta_{r_k}(z^k) = \theta\left(\frac{z^k}{r_k}\right),$$

and so, computing the derivative gives

$$t_k \theta' r_k(z^k) = \frac{t_k}{r_k} \theta'\left(\frac{z^k}{r_k}\right).$$

So, either $z^k = o(r_k)$ and by $0 < \theta'(0) < \infty$

$$\lim_{k\to\infty} t_k \theta'_{r_k}(z_k) = \lim_{k\to\infty} \frac{t_k}{r_k} \theta'\left(\frac{z^k}{r_k}\right) = \lim_{k\to\infty} \frac{t_k}{r_k} \theta'(0).$$

Either there exists a constant $C > 0$ such that $z^k = C r_k$ and so

$$\lim_{k\to\infty} t_k \theta'_{r_k}(z_k) = \lim_{k\to\infty} \frac{t_k}{r_k} \theta'\left(\frac{C r^k}{r_k}\right) = \lim_{k\to\infty} \frac{t_k}{r_k} \theta'(C).$$

Otherwise for k sufficiently large $r_k \leq z_k$ and by concavity of θ_r

$$0 \leq \lim_{k\to\infty} t_k \theta'_{r_k}(z_k) \leq \lim_{k\to\infty} t_k \theta'_{r_k}(r_k) = \lim_{k\to\infty} \frac{t_k}{r_k} \theta'(1).$$

□

The following lemma is a direct application of Lemma 3.2 on the convergence of the butterfly relaxation when t and r go to zero.

Lemma 3.3 (Convergence of $\Phi^B(G(x), H(x); \hat{t})$)**.** *Assume that* (MPCC) *has a non-empty feasible set. Given two sequences* $\{r_k\}$ *and* $\{t_k\}$*, which converge to* 0 *as* k *goes to infinity and* $\forall k \in \mathbb{N},\ (r_k, t_k) \in \mathbb{R}^2_{++}$*. Let* $\{x^k\}$ *be a sequence of points such that* $\lim_{k\to\infty} x^k = x^*$ *and satisfying for all* $i \in \{1, \ldots, q\}$ *and for all* $k \in \mathbb{N}$

$$G_i(x^k) \geq -\bar{r},\ H_i(x^k) \geq -\bar{r},\ \Phi^B_i(G(x), H(x); \hat{t})(x^k) \leq 0.$$

Then, x^* *is a feasible point for* (MPCC) *as long as* $g(x^*) \leq 0$ *and* $h(x^*) = 0$*.*

We conclude this section by an example that shows that the butterfly relaxation may improve relaxations from [46] and [48]. Indeed, it illustrates an example where there are no sequences of stationary points that converge to some undesirable point.

Example 3.1.

$$\min_{x\in\mathbb{R}^2} -x_1 \ s.t\ x_1 \leq 1,\ 0 \leq x_1 \perp x_2 \geq 0.$$

In this example, there are two stationary points : an S-stationary point $(1,0)$ *that is the global minimum and an M-stationary point* $(0,0)$*. Since, obviously MPCC-LICQ holds at the latter point it is undesirable by Theorem 2.2.*

Unlike the relaxations (KDB) *and* (KS) *where for* $t_k = \frac{1}{k}$ *a sequence* $x^k = (t_k\ t_k)^T$ *may converge to* $(0,0)$ *as* k *goes to infinity there is no sequences of stationary point that converges to this undesirable point with the butterfly relaxation.*

4 Theoretical Properties

The study of theoretical properties of the butterfly relaxation methods is split into three parts: convergence of the sequence of stationary points, existence of Lagrange multipliers for the relaxed non-linear program and convergence of the sequence of ϵ-stationary points. In each case, this new technique discussed here is proved to get the best-known theoretical properties for some choices of the parameters.

4.1 Convergence

In this section, we focus on the convergence properties of the butterfly relaxation methods and the constraint qualifications guaranteeing convergence of the sequence of stationary points generated by the methods. Our aim is to compute an M-stationary point or at least to provide a certificate if we converge to an undesirable point.

Relaxation methods that converge to M-stationary points are introduced in [46] and [48]. C-stationary points are also frequent guests in these relaxations methods as in [65] and [54].

We prove in Theorem 4.1 that butterfly relaxations converge to an A-stationary point and provide a certificate independent of the multipliers in the case it converges to undesirable points. This result is improved to a convergence to M-stationary points for some choices on the parameters t and r in Corollary 4.1.

Another main concern in the literature is to find the weakest constraint qualification, which ensures convergence of the method. This has been extensively studied in the thesis [67] and related papers mentioned herein, where they prove convergence of most of the existing relaxation methods in the literature under a hypothesis close to MPCC-CRCQ. More recently in [63] the author proves convergence of the relaxation from [46] and [48] under MPCC-CCP.

Convergence of the butterfly relaxations under MPCC-CRCQ is proved in Proposition 4.1. An improved result for some choices of the parameter t and r is given in Proposition 4.2 that uses our new constraint qualification denoted MPCC-CRSC. Example 4.4 shows that our methods may not converge under MPCC-CCP since it requires boundedness of some multipliers.

Theorem 4.1. *Given two sequences* $\{t_k\}$ *and* $\{r_k\}$ *decreasing to zero such that* $(t_k, r_k) \in \mathbb{R}^2_+$ *satisfying* (6)*. Let* $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ *be a sequence of points from* $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{3q}$ *that are stationary points of* $R^B_{\hat{t}}(x^k)$ *for all* $k \in \mathbb{N}$ *with* $x^k \to x^*$*. Assume that the sequence*

$$\{\lambda^{g,k}, \lambda^{h,k}, \eta^{G,k}, \eta^{H,k}\}, \tag{8}$$

where for all $i \in \{1, \ldots, q\}$

$$\eta_i^{G,k} := \lambda^{G,k} - \lambda_i^{\Phi,k}\left(F_{1i}(x^k;\hat{t}_k) - t_k\theta'_{r_k}\left(G_i(x^k)\right)F_{2i}(x^k;\hat{t}_k)\right),$$
$$\eta_i^{H,k} := \lambda^{H,k} - \lambda_i^{\Phi,k}\left(F_{2i}(x^k;\hat{t}_k) - t_k\theta'_{r_k}\left(H_i(x^k)\right)F_{1i}(x^k;\hat{t}_k)\right),$$

is bounded. Then, one of the three following case holds:

(i) x^* *is an S-stationary point, if for all* $i \in \mathcal{I}^{00}(x^*)$ *one of the following holds:*

(i.a) there exists $\bar{k} \in \mathbb{N}$ *such that* $G_i(x^k) = H_i(x^k) = -\bar{r}_k,\ \forall k \geq \bar{k}$;

(i.b) the sequence of multiplier $\{\lambda_i^{\Phi,k}\}$ *is bounded;*

(i.c) $\{\lambda_i^{\Phi,k}\}$ *is unbounded with* $\lim_{k\to\infty} \lambda_i^{\Phi,k} F_{2i}(x^k;\hat{t}_k) = 0$ *for* $i \in \mathcal{I}_{GH}^{0+}(x^k;\hat{t}_k)$ *and* $\lim_{k\to\infty} \lambda_i^{\Phi,k} F_{1i}(x^k;\hat{t}_k) = 0$ *for* $i \in \mathcal{I}_{GH}^{+0}(x^k;\hat{t}_k)$.

(ii) x^* *is an M-stationary point, if for all* $i \in \mathcal{I}^{00}(x^*)$ *that do not satisfy conditions (i) the sequence of multiplier* $\{\lambda_i^{\Phi,k}\}$ *is unbounded and either for* $i \in \mathcal{I}_{GH}^{0+}(x^k;\hat{t}_k)$, $\lim_{k\to\infty} t_k \theta'_{r_k}(G_i(x^k)) = 0$ *either for* $i \in \mathcal{I}_{GH}^{+0}(x^k;\hat{t}_k)$, $\lim_{k\to\infty} t_k \theta'_{r_k}(H_i(x^k)) = 0$.

(iii) x^* *is an A-stationary point, if the sequence of multiplier diverges or for all* $i \in \mathcal{I}^{00}(x^*)$ *that do not satisfy conditions (ii) the sequence of multiplier* $\{\lambda_i^{\Phi,k}\}$ *is unbounded and either for all* $i \in \mathcal{I}_{GH}^{0+}(x^k;\hat{t}_k)$, $\lim_{k\to\infty} t_k \theta'_{r_k}(G_i(x^k)) > 0$ *either for all* $i \in \mathcal{I}_{GH}^{+0}(x^k;\hat{t}_k)$, $\lim_{k\to\infty} t_k \theta'_{r_k}(H_i(x^k)) > 0$.

The boundedness assumption on the sequence (8) is a classical assumption and is guaranteed under some constraint qualification as shown in the next Proposition 4.1.

Proof. First, we identify the expressions of the multipliers of the complementarity constraint in Definition 2.7 in function of the stationary points of $R_{\hat{t}}^B(x^k)$. Let $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ be a sequence of KKT points of $R_{\hat{t}}^B$ for all $k \in \mathbb{N}$, that by definition satisfies

$$\begin{aligned} 0 = \nabla f(x^k) &+ \sum_{i=1}^{p} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) - \sum_{i=1}^{q} \lambda_i^{G,k} \nabla G_i(x^k) \\ &- \sum_{i=1}^{q} \lambda_i^{H,k} \nabla H_i(x^k) + \sum_{i=1}^{q} \lambda_i^{\Phi,k} \nabla \Phi_i^B(G(x^k), H(x^k); \hat{t}_k), \end{aligned}$$

with

$$\begin{aligned} &\lambda_i^{g,k} = 0,\ \forall i \notin \mathcal{I}_g(x^k) \text{ and } \lambda_i^{g,k} \geq 0,\ \forall i \in \mathcal{I}_g(x^k) \\ &\lambda_i^{G,k} = 0,\ \forall i \notin \mathcal{I}_G(x^k) \text{ and } \lambda_i^{G,k} \geq 0,\ \forall i \in \mathcal{I}_G(x^k) \\ &\lambda_i^{H,k} = 0,\ \forall i \notin \mathcal{I}_H(x^k) \text{ and } \lambda_i^{H,k} \geq 0,\ \forall i \in \mathcal{I}_H(x^k) \\ &\lambda_i^{\Phi,k} = 0,\ \forall i \notin \mathcal{I}_{GH}(x^k;\hat{t}_k) \text{ and } \lambda_i^{\Phi,k} \geq 0,\ \forall i \in \mathcal{I}_{GH}(x^k;\hat{t}_k). \end{aligned}$$

Since the representation of $\Phi_i^B(G(x), H(x); \hat{t})(x^k)$ immediately gives $\nabla \Phi_i^B(G(x), H(x); \hat{t})(x^k) = 0,\ \forall i \in \mathcal{I}_{GH}^{00}(x^k;\hat{t}_k)$ for all $k \in \mathbb{N}$. Thus, we can rewrite the equation above as

$$-\nabla f(x^k) = \sum_{i=1}^{p} \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) - \sum_{i=1}^{q} \eta_i^{G,k} \nabla G_i(x^k) - \sum_{i=1}^{q} \eta_i^{H,k} \nabla H_i(x^k), \tag{9}$$

where

$$\eta_i^{G,k} = \begin{cases} \lambda^{G,k}, & \text{if } i \in \mathcal{I}_G(x^k;\hat{t}_k) \\ -\lambda_i^{\Phi,k} t_k \theta'_{r_k}(G_i(x^k)) F_{2i}(x^k;\hat{t}_k), & \text{if } i \in \mathcal{I}_{GH}^{0+}(x^k;\hat{t}_k) \\ \lambda_i^{\Phi,k} F_{1i}(x^k;\hat{t}_k), & \text{if } i \in \mathcal{I}_{GH}^{+0}(x^k;\hat{t}_k) \\ 0, & \text{otherwise,} \end{cases}$$

$$\eta_i^{H,k} = \begin{cases} \lambda^{H,k}, & \text{if } i \in \mathcal{I}_H(x^k;\hat{t}_k) \\ -\lambda_i^{\Phi,k} t_k \theta'_{r_k}(H_i(x^k)) F_{1i}(x^k;\hat{t}_k), & \text{if } i \in \mathcal{I}_{GH}^{+0}(x^k;\hat{t}_k) \\ \lambda_i^{\Phi,k} F_{2i}(x^k;\hat{t}_k), & \text{if } i \in \mathcal{I}_{GH}^{0+}(x^k;\hat{t}_k) \\ 0, & \text{otherwise.} \end{cases}$$

Noticing that whenever $i \in \{i = 1, \dots, q \mid F_1(x; r, t)_i(x^k; \hat{t}_k) = 0\}$ implies that $i \in \mathcal{I}^{0+}_{GH}(x^k; \hat{t}_k)$ or symmetrically $i \in \{i \mid F_2(x; t, r)_i(x^k, \hat{t}_k) = 0\}$ implies that $i \in \mathcal{I}^{+0}_{GH}(x^k; \hat{t}_k)$ by concavity and $t_k\theta'(0) \leq r_k$ for all $k \in \mathbb{N}$.

Since we admit that the sequence $\{\lambda^{g,k}, \lambda^{h,k}, \eta^{G,k}, \eta^{H,k}\}$ is bounded and then converges, up to a subsequence, to some limit denoted by $\{\lambda^{g,*}, \lambda^{h,*}, \eta^{G,*}, \eta^{H,*}\}$.

These multipliers are well-defined since

$$\mathcal{I}_G(x^k; \hat{t}_k) \cap \mathcal{I}_{GH}(x^k; \hat{t}_k) \cap \big(\{1, \dots, q\} \setminus \mathcal{I}^{00}_{GH}(x^k; \hat{t}_k)\big) = \emptyset$$
$$\mathcal{I}_H(x^k; \hat{t}_k) \cap \mathcal{I}_{GH}(x^k; \hat{t}_k) \cap \big(\{1, \dots, q\} \setminus \mathcal{I}^{00}_{GH}(x^k; \hat{t}_k)\big) = \emptyset,$$

and for k sufficiently large

$$\begin{aligned}
&\operatorname{supp}(\lambda^{G,k}) \subseteq \mathcal{I}_G(x^k; \hat{t}_k)\\
&\operatorname{supp}(\lambda^{H,k}) \subseteq \mathcal{I}_H(x^k; \hat{t}_k)\\
&\operatorname{supp}(\lambda^{G,k}) \subseteq \mathcal{I}_{GH}(x^k; \hat{t}_k)\\
&\operatorname{supp}(\eta^{G,k}) \subseteq \mathcal{I}_{GH}(x^k; \hat{t}_k) \cap (\{1, \dots, q\} \setminus \mathcal{I}^{00}_{GH}(x^k; \hat{t}_k))\\
&\operatorname{supp}(\eta^{H,k}) \subseteq \mathcal{I}_{GH}(x^k; \hat{t}_k) \cap (\{1, \dots, q\} \setminus \mathcal{I}^{00}_{GH}(x^k; \hat{t}_k)).
\end{aligned}$$

Moreover, for k sufficiently large it holds

$$\begin{aligned}
&\operatorname{supp}(\lambda^{G,*}) \subseteq \operatorname{supp}(\lambda^{G,k})\\
&\operatorname{supp}(\lambda^{H,*}) \subseteq \operatorname{supp}(\lambda^{H,k})\\
&\operatorname{supp}(\eta^{G,*}) \subseteq \operatorname{supp}(\eta^{G,k})\\
&\operatorname{supp}(\eta^{H,*}) \subseteq \operatorname{supp}(\eta^{H,k}).
\end{aligned}$$

Proof that shows convergence of the sequence and weak-stationarity of x^* is given by Lemma 4.3 for $\epsilon_k = 0$.

So, let us verify that for i in $\mathcal{I}^{00}$ in some cases x^* is an M- or an A-stationary point. Consider the various possible cases, where we denote

$$\lambda^G_0 := \{i = 1, \dots, q \mid \lambda^{G,*}_i = \eta^{G,*}_i\} \text{ and } \lambda^H_0 := \{i = 1, \dots, q \mid \lambda^{H,*}_i = \eta^{H,*}_i\} :$$

1. If $i \in \operatorname{supp}(\lambda^{G,*}) \cap \operatorname{supp}(\lambda^{H,*})$, then for k sufficiently large $i \in \operatorname{supp}(\lambda^{G,k}) \cap \operatorname{supp}(\lambda^{H,k})$. One has $\lambda^{G,k}_i \geq 0$, $\lambda^{H,k}_i \geq 0$ and
$$G_i(x^k) = H_i(x^k) = -\bar{r}_k.$$

2. If $i \in \operatorname{supp}(\lambda^{G,*}) \cap \operatorname{supp}(\eta^{H,*})$, then for k sufficiently large $i \in \operatorname{supp}(\lambda^{G,k}) \cap \operatorname{supp}(\eta^{H,k})$. One has $\lambda^{G,k}_i \geq 0$, $G_i(x^k) = -\bar{r}_k$ and necessarily $i \in \mathcal{I}_{GH}(x^k; \hat{t}_k)$, which is not possible.

3. The case $i \in \operatorname{supp}(\eta^{G,*}) \cap \operatorname{supp}(\lambda^{H,*})$ is completely similar.

4. If $i \in \operatorname{supp}(\lambda^{G,*}) \cap \lambda^H_0$, then $\eta^{G,*}_i \geq 0$ and $\eta^{H,*}_i = 0$.

5. If $i \in \lambda^G_0 \cap \operatorname{supp}(\lambda^{H,*})$, then $\eta^{G,*}_i \geq 0$ and $\eta^{H,*}_i = 0$.

6. If $i \in \lambda^G_0 \cap \lambda^H_0$, then $\eta^{G,*}_i = \eta^{H,*}_i = 0$.

7. If $i \in \lambda^G_0 \cap \operatorname{supp}(\eta^{H,*})$, then $i \in \lambda^G_0 \cap \operatorname{supp}(\eta^{H,k})$. Since $\eta^{G,k} = 0$ and $\eta^{H,k}$ free, one has $\lambda^{\Phi,k}_i \geq 0$ and then $i \in \mathcal{I}_{GH}(x^k; \hat{t}_k)$.
$$\eta^{G,k}_i = 0 \Longleftrightarrow F_{1i}(x^k; \hat{t}_k) = t_k\theta'_{r_k}(G_i(x^k))F_{2i}(x^k; \hat{t}_k) \text{ or } \lambda^{\Phi,k}_i = 0.$$
Moreover $t_k\theta'_{r_k}(G_i(x^k)) > 0$, so either $\lambda^{\Phi,k}_i = 0$ or $F_{1i}(x^k; \hat{t}_k) = F_{2i}(x^k; \hat{t}_k) = 0$. It follows that $\eta^{G,*}_i = \eta^{H,*}_i = 0$.

8. The case $i \in \text{supp}(\eta^{G,*}) \cap \lambda_0^H$ is completely similar to the previous case and leads to $\hat{\gamma} = \hat{\nu} = 0$.

9. If $i \in \text{supp}(\eta^{G,*}) \cap \text{supp}(\eta^{H,*})$, then $i \in \text{supp}(\eta^{G,k}) \cap \text{supp}(\eta^{H,k})$ for k sufficiently large and $i \in \mathcal{I}_{GH}(x^k; \hat{t}_k)$.

   (a). $i \in \mathcal{I}_{GH}^{00}(x^k; \hat{t}_k)$ implies that $F_{1i}(x^k; \hat{t}_k) = F_{2i}(x^k; \hat{t}_k)$, therefore $G(x^k) = H(x^k) = 0$ and $\eta_i^{G,*} = \eta_i^{H,*} = 0$.

   (b). If $i \in \mathcal{I}_{GH}^{0+}(x^k; \hat{t}_k)$, then $F_{1i}(x^k; \hat{t}_k) = 0$

$$0 < H_i(x^k) = t_k \theta_{r_k}(G_i(x^k)) < \frac{t_k \theta'(0)}{r_k} G_i(x^k),$$

   therefore $F_{2i}(x^k; \hat{t}_k) > 0$. Assume $\lambda_i^{\Phi,k}$ is not bounded, then going through the limit there is a non-negative constant C such that

$$\lim_{k \to \infty} \lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k) = C \geq 0,$$

   and so $\eta_i^{H,*} = -C$. If $\lambda_i^{\Phi,k}$ is bounded, it corresponds to the case $C = 0$. Furthermore either one has

$$\lim_{k \to \infty} t_k \theta'_{r_k}(G_i(x^k)) \geq 0$$

   and so $\eta_i^{G,*} \geq 0$ and $\eta_i^{H,*} \leq 0$. Either one has

$$\lim_{k \to \infty} t_k \theta'_{r_k}(G_i(x^k)) = 0$$

   and so $\eta_i^{G,*} = 0$ and $\eta_i^{H,*} < 0$.

   (c). The case $i \in \mathcal{I}_{GH}^{+0}(x^k; \hat{t}_k)$ is completely similar to the previous case.

Indices that correspond to the first eight cases and 9.a) are indices that satisfy S-stationary condition. Furthermore, the indices in cases 9.b) and 9.c), when the constant $C = 0$, also have the sign of S-stationary indices.

M- and A-stationary indices may appear only in the case 9.b) when $C \neq 0$ and either $t_k \theta'_{r_k}(G_i(x^k)) = 0$ or $t_k \theta'_{r_k}(G_i(x^k)) > 0$ for $i \in \mathcal{I}_{GH}^{+0}(x^k; \hat{t}_k)$ and symmetrically in case 9.c). □

The following proposition proves convergence of the sequence of multipliers under MPCC-CRCQ by a direct application of Theorem 2.3. Indeed, assumption (3) of Theorem 2.3 is guaranteed by Theorem 4.1.

Proposition 4.1. *Given two sequences $\{t_k\}$ and $\{r_k\}$ satisfying (6) such that $\forall k \in \mathbb{N}, (t_k, r_k) \in \mathbb{R}_+^2$, both sequences decrease to zero as k goes to infinity. Let $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ be a sequence of points that are stationary points of $R_{\hat{t}}^B(x^k)$ for all $k \in \mathbb{N}$ with $x^k \to x^*$ such that MPCC-CRCQ holds at x^*. Then, the sequence (8) is bounded.*

In [63], the author proves similar convergence results for the relaxations [46] and [48] using the very weak constraint qualification MPCC-CCP, obtained by deriving the sequential optimality conditions from [9] in non-linear programming to (MPCC). However, this constraint qualification does not ensure boundedness of the sequence of multipliers (8), which is necessary for our previous theorem. The following example shows that the result of Proposition 4.1 is sharp since convergence can not be ensured if MPCC-MFCQ holds at the limit point. Furthermore, since MPCC-MFCQ is stronger than MPCC-CRSC a similar behaviour can be observed with the later MPCC-CQ.

Example 4.1. *Consider the following two dimensional example*

$$\min_{(x_1, x_2) \in \mathbb{R}^2} x_2 \ s.t \ 0 \leq x_1 + x_2^2 \perp x_1 \geq 0.$$

MPCC-MFCQ holds at $(0,0)^T$. *However MPCC-CRCQ obviously fails to hold at this point. In this case the point* $(0,0)^T$ *is even not a weak-stationary point. Indeed, given a sequence* $x^k \in \mathcal{I}_{GH}$ *such that* $x^k \to (0,0)^T$ *then* $\lambda^{G,k} = \lambda^{H,k} = 0$ *and we can choose* $\lambda^{\Phi,k}$ *that satisfies*

$$\eta^{G,k} = -\eta^{H,k} = \frac{1}{2x_2^k}.$$

So, in this case there exists a sequence of stationary points that converges to an highly undesirable point.

Theorem 4.1 describes the various sequences that can arise from these relaxation methods in a constructive way. Indeed, it shows that in general the butterfly relaxations may converge to some undesirable A-stationary points. This theorem also provides a certificate independent of the computed multiplier that detects during the iterations, whether the method converges to this kind of undesirable point.

According to condition (iii) from Theorem 4.1 if we detect for k sufficiently large that there exists an index $i \in \mathcal{I}^{00}(x^*)$ such that

$$\text{either } i \in \mathcal{I}_{GH}^{0+}(x^k; \hat{t}_k),\ \lim_{k\to\infty} t_k \theta'_{r_k}(G_i(x^k)) > 0 \text{ either } i \in \mathcal{I}_{GH}^{+0}(x^k; \hat{t}_k),\ \lim_{k\to\infty} t_k \theta'_{r_k}(H_i(x^k)) > \ 0$$

then x^k converges to an A-stationary point and not more. This is a priori not a trivial task, since the set of multiplier at those points is not bounded and an M-stationary point may be defined for only a subset of multipliers among the unbounded set of multiplier at this point. The following examples illustrate this phenomenon.

Example 4.2.

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_1^2 + x_1x_2 + x_2^2 - \frac{1}{2}x_1 + x_2$$
$$0 \le x_1 \perp x_2 \ge 0$$

There exist two stationary points : an A-stationary point $(0,0)^T$ *and the global minimum* $(\frac{1}{4},0)^T$. *Similar computation gives two stationary points for the Butterfly relaxation* (5) *with* $t_k = r_k,\ \forall k \in \mathbb{N}$ *:* $(1/2,0)^T$ *and a point* $(x_1,x_2)^T$ *such that*

$$x_1 = t\theta_r(x_2),\ 0 = \frac{1}{2} + t\theta_r(x_2) + 2x_2 - t\theta'_r(x_2)(1 - 2t\theta_r(x_2) - x_2).$$

Example 4.3 gives another example with an A-stationary point. However, in this case the butterfly relaxation can not converges to this point as the multiplier are too large.

Example 4.3. *This example is very similar to the previous example and consider*

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_1^2 + x_1x_2 + x_2^2 - x_1 + x_2$$
$$0 \le x_1 \perp x_2 \ge 0$$

There exists two stationary point : an A-stationary point $(0,0)^T$ *with multipliers* $\lambda^G = -1$, $\lambda^H = 1$ *and the global minimum* $(\frac{1}{2},0)^T$ *with multipliers* $\lambda^G = 0,\ \lambda^H = \frac{3}{2}$, *which is S-stationary. Notice that* λ^G *and* λ^H *are unique for each stationary point since MPCC-LICQ holds at any feasible point.*

Simple computation gives two stationary points for the Butterfly relaxation (5) *with* $t_k = r_k,\ \forall k \in \mathbb{N}$ *:* $(1/2,0)^T$ *and a point* (x,y) *such that*

$$x_1 = t\theta_r(x_2),\ 0 = 1 + t\theta_r(x_2) + 2x_2 - t\theta'_r(x_2)(1 - 2t\theta_r(x_2) - x_2).$$

However, the previous equations never holds for $t\theta'(0) \le r$.

The following corollary of Theorem 4.1 shows that for some choice of parameters we can get rid of the undesirable A-stationary points. It is an essential result, since it shows that a subfamily of the butterfly relaxations has the desired convergence property.

Corollary 4.1. *Given two sequences $\{t_k\}$ and $\{r_k\}$ decreasing to zero such that $(t_k, r_k) \in \mathbb{R}^2_+ \ \forall k \in \mathbb{N}$ and $t_k = o(r_k)$ for k sufficiently large. Let $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ be a sequence of points that are stationary points of $R^B_{\hat{t}}(x^k)$ for all $k \in \mathbb{N}$ with $x^k \to x^*$. Assume that the sequence (8) is bounded. Then, x^* is an M-stationary point.*

Proof. Condition (iii) Theorem 4.1 can not hold by Lemma 3.2. □

For this relation between the parameters t and r, we can improve the result of Proposition 4.1 by a straightforward application of Corollary 2.2.

Proposition 4.2. *Given two sequences $\{t_k\}$ and $\{r_k\}$ decreasing to zero such that $(t_k, r_k) \in \mathbb{R}^2_+ \ \forall k \in \mathbb{N}$ and $t_k = o(r_k)$ for k sufficiently large. Let $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ be a sequence of points that are stationary points of $R^B_{\hat{t}}(x^k)$ for all $k \in \mathbb{N}$ with $x^k \to x^*$ such that MPCC-CRSC holds at x^*. Then, the sequence (8) is bounded.*

In conclusion, a sequence of stationary points of the butterfly relaxation with $t_k = o(r_k)$ that satisfies MPCC-CRSC at its limit point converges to an M-stationary point.

The following example shows that this result is sharp, since it illustrates an example where MPCC-CRSC does not hold and the method converge to a undesirable weak-stationary point. This phenomenon only happen if the sequence of multipliers (8) is unbounded multiplier.

Example 4.4.

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_2^2 \ s.t\ 0 \le x_1^2 \perp x_1 + x_2^2 \ge 0.$$

The feasible set of this example is the set $\mathcal{Z} = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1 = 0\} \cup \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1 = -x_2^2\}$. There is a unique stationary point $(0, 0)$, which is M-stationary with $(\lambda^G, \lambda^H = 0)$.

It is easy to verify that MPCC-CCP holds at this point. However, MPCC-CRSC fails to hold at any point $(0, a \in \mathbb{R})^T$ since the gradient of x_1^2 is non-zero for $x \neq 0$.

In this example the butterfly relaxation method may fail to converge to a weak-stationary point. Indeed, for $x^* = (0, a \neq 0)^T$ we can find a sequence x^k such that for t_k, r_k sufficiently small $F_2(x^k; \hat{t}_k) = 0$ and

$$x_1^k = t_k \theta'_{r_k}(x_1^k + a^2),\ x_2^k = a,\ \lambda^{\Phi,k} F_1(x^k; \hat{t}_k) = \frac{1}{-t_k \theta'_{r_k}(x_1^k + a^2)}$$

In this case, we have

$$\eta^{G,k} = \frac{1}{t_k \theta'_{r_k}(x_1^k + a^2)} \to \infty \text{ and } \eta^{H,k} = -1,$$

which is not a weak stationary point, since $\eta^{H,k} \neq 0$.

We conclude this section by the case where (MPCC) satisfies MPCC-LICQ and then M-stationary point are undesirable, as for instance in Example 4.5.

It can be seen from their definitions that M- and S-stationary points coincide if strict complementarity holds. That is why in the literature convergence to an S-stationary point is usually guaranteed under an hypothesis relative of the indices that do not satisfy strict complementarity. An asymptotically weakly non-degenerate hypothesis was first introduced in [28] in this context.

Definition 4.1. *A sequence $\{x^k\}$ is asymptotically weakly non-degenerate, if $x^k \to x^*$ as $\{\hat{t}_k\} \downarrow 0$ and there is a $\hat{t}^*$ such that for $t \in (0, \hat{t}^*)$ one has*

$$-1 \le \frac{G_i(x^k)}{H_i(x^k)} \le 1,\ (i \in \mathcal{I}^{00}(x^*) \mid H_i(x^k) \neq 0) \text{ and } -1 \le \frac{H_i(x^k)}{G_i(x^k)} \le 1,\ (i \in \mathcal{I}^{00}(x^*) \mid G_i(x^k) \neq 0)$$

In [46], convergence of their approximation scheme to an S-stationary point has been proved if the sequence $\{x^k\}$ satisfies the asymptotically weakly non-degenerate assumption for (KDB). Theorem 4.1 gives a full description of the sequences of points converging to an S- stationary point and a similar result could be derived here. This is illustrated by the following example.

Example 4.5.

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_1^2 - x_1x_2 + \frac{1}{3}x_2^2 - 2x_1$$
$$0 \le x_1 \perp x_2 \ge 0$$

There is one S-stationary point $(1,0)^T$ *and one M-stationary point* $(0,0)^T$*. It is essential to note here that MPCC-LICQ holds at* $(0,0)^T$ *and so this point is undesirable.*

Using the relaxation (KDB) *we get two M-stationary points* $(1,0)^T\ \forall t$ *and* $(t,\frac{3}{2}t)^T$ *which doesn't satisfy asymptotically weakly non-degenerate assumption.*

Considering the butterfly relaxation, it follows that $(x_1,x_2,\lambda^G,\lambda^H,\lambda^\Phi)\in\mathbb{R}^2\times\mathbb{R}^3_+$ *satisfy*

$$0 = 2x_1 - x_2 - 2 - \lambda^G + \lambda^\Phi(F_1(x;\hat{t}) - t\theta_r'(x_1)F_2(x;\hat{t}))$$
$$0 = -x_1 + \frac{2}{3}x_2 - \lambda^H + \lambda^\Phi(F_2(x;\hat{t}) - t\theta_r'(x_2)F_1(x;\hat{t})),$$

with $\lambda^\Phi F_1(x;\hat{t})F_2(x;\hat{t}) = 0$, $\lambda^G x_1 = 0$, $\lambda^H x_2 = 0$*. There exists a sequence* $(x^k_{1,S},x^k_{2,S})^T$ *that tends to* $(1,0)^T$*, which verifies*

$$x_1^k = t_k\theta_{r_k}(x_1^k),\ 2x_1^k - x_2^k - 2 - t_k\theta'_{r_k}(x_1^k)(x_2^k - \frac{2}{3}x_1^k) = 0,\ \text{with } \lambda_S^{\Phi,k}\to 1,$$

and a sequence $(x^k_{1,M},x^k_{2,M})^T$ *that tends to* $(0,0)^T$ *, which verifies*

$$x_1^k = t_k\theta_{r_k}(x_2^k),\ x_1^k - \frac{2}{3}x_2^k - t_k\theta'_{r_k}(x_2^k)(2 - 2x_1^k + x_2^k) = 0,\ \text{with } x_1^k = \omega(r_k),\ \lambda_M^{\Phi,k}\to\infty.$$

However, the latter sequence does not satisfy the asymptotically weakly non-degenerate hypothesis for t *and* r *sufficiently small.*

It should be noted however that this result is not sharp as shown by the following example.

Example 4.6 (Tangi)**.**

$$\min_{(x_1,x_2)^T\in\mathbb{R}^2} (2x_1 - x_2)^4 + \frac{1}{x_1x_2+1}\ s.t\ 0 \le x_1 \perp x_2 \ge 0.$$

It is clear that $(0,0)^T$ *is the global minimum and satisfies MPCC-LICQ. So, it is an S-stationary point by Theorem 2.2. The relaxation* (KDB) *possesses a global minimum in* $x(t) = (\frac{t}{2},t)^T$*, which converges to* $(0,0)^T$ *as* t *goes to 0 and does not satisfy the asymptotically weakly non-degenerate assumption.*

According to [65] one important feature of stationary point that satisfies MPCC-LICQ is that the multiplier associated are unique. Further research may propose heuristics to try to avoid this undesirable case of computing an M-stationary point that satisfies MPCC-LICQ.

4.2 Existence of Lagrange Multipliers of the Relaxed sub-problems

In this part, we consider regularity properties of the relaxed non-linear programs. Indeed, in order to guarantee the existence of a sequence of stationary points the relaxed non-linear programs must satisfy some constraint qualifications in the neighborhood of the limit point. The butterfly relaxations satisfy Guignard CQ as stated in Theorem 4.2, which is equivalent in term of regularity to the relaxation (KS). The butterfly relaxations with $t\theta'(0) = r$ are more regular as they satisfy Abadie CQ, see Theorem 4.3.

In our proofs we use the following results from [67] which allows to compute the tangent cone of $\mathcal{X}_t^B$ and its polar.

Lemma 4.1. *[[67], Lemma 8.10] For all $t > 0$ and all x feasible for $R_{\hat{t}}^B$,*

$$\mathcal{T}_{\mathcal{X}_{\hat{t}}^B}(x) = \cup_{I \subseteq \mathcal{I}_{GH}^{00}(x;\hat{t})} \mathcal{T}_{\mathcal{X}(\hat{t},I)}(x),$$

where $\mathcal{X}(\hat{t}, I)$ is the feasible set of the non-linear program $NLP_{t,I}(x)$ with $I \subseteq \mathcal{I}_{GH}^{00}(x;\hat{t})$ defined as

$$\begin{aligned}
&\min_{x \in \mathbb{R}^n} f(x) \\
&g(x) \leq 0,\ h(x) = 0, \\
&G(x) \geq -\bar{r},\ H(x) \geq -\bar{r}, \\
&\Phi_i^B(G(x), H(x); \hat{t}) \leq 0,\ i \notin \mathcal{I}_{GH}^{00}(x;\hat{t}), \\
&F_{1i}(x;\hat{t}) \leq 0,\ F_{2i}(x;\hat{t}) \geq 0,\ i \in I, \\
&F_{1i}(x;\hat{t}) \geq 0,\ F_{2i}(x;\hat{t}) \leq 0,\ i \in I^c,
\end{aligned} \tag{NLP$_{t,I}(x)$}$$

where $I \cup I^c = \mathcal{I}_{GH}^{00}(x;\hat{t})$ and $I \cap I^c = \emptyset$.

We also need the following lemma that links the gradients of G and H with the gradients of $F_1(x; r, t)$ and $F_2(x; t, r)$.

Lemma 4.2. *Let $I \in \mathcal{P}(\mathcal{I}_{GH}^{00}(x;\hat{t}))$. Assume that the gradients*

$$\begin{aligned}
&\{\nabla g_i(x)\ (i \in \mathcal{I}_g(x)),\ \nabla h_i(x)\ (i = 1, \ldots, m), \\
&\nabla G_i(x)\ (i \in \mathcal{I}_{GH}^{00}(x;\hat{t}) \cup \mathcal{I}_{GH}^{+0}(x;\hat{t})),\ \nabla H_i(x^*)\ (i \in \mathcal{I}_{GH}^{00}(x;\hat{t}) \cup \mathcal{I}_{GH}^{0+}(x;\hat{t}))\}
\end{aligned}$$

are linearly independent. Then, MFCQ holds at x for (NLP$_{t,I}(x)$).

Proof. We show that the gradients of the constraints of (NLP$_{t,I}(x)$) are positively linearly independent. For this purpose, we prove that the trivial solution is the only solution to the equation

$$\begin{aligned}
0 = &\sum_{i \in \mathcal{I}_g(x)} \lambda_i^g \nabla g_i(x) + \sum_{i=1}^m \lambda_i^h \nabla h_i(x) + \sum_{i \in \mathcal{I}_G(x;\hat{t})} \lambda_i^G \nabla G_i(x) + \sum_{i \in \mathcal{I}_H(x;\hat{t})} \lambda_i^H \nabla H_i(x) \\
+ &\sum_{i \in \mathcal{I}_{GH}^{+0}(x;\hat{t}) \cup \mathcal{I}_{GH}^{0+}(x;\hat{t})} \lambda_i^\Phi \nabla \Phi_i^B(G(x), H(x); \hat{t}) + \sum_{i \in I} \nu_i^{F_1} \nabla F_{1i}(x;\hat{t}) - \sum_{i \in I} \nu_i^{F_2} \nabla F_{2i}(x;\hat{t}) \\
&- \sum_{i \in I^c} \mu_i^{F_1} \nabla F_{1i}(x;\hat{t}) + \sum_{i \in I^c} \mu_i^{F_2} \nabla F_{2i}(x;\hat{t}),
\end{aligned}$$

with $\lambda^g, \lambda^G, \lambda^H, \lambda^\Phi, \nu^{F_1}, \nu^{F_2}, \mu^{F_1}, \mu^{F_2} \geq 0$. By definition of $F_1(x; r, t)$ and $F_2(x; t, r)$ it holds that

$$\begin{aligned}
\nabla F_{1i}(x;\hat{t}) &= \nabla H_i(x) - t\theta_r'(G_i(x))\nabla G_i(x) \\
\nabla F_{2i}(x;\hat{t}) &= \nabla G_i(x) - t\theta_r'(H_i(x))\nabla H_i(x).
\end{aligned}$$

The gradients of $\Phi^B(G(x), H(x); \hat{t})$ is given by Lemma 3.1.

We now replace those gradients in the equation above

$$\begin{aligned}
0 = &\sum_{i \in \mathcal{I}_g(x)} \lambda_i^g \nabla g_i(x) + \sum_{i=1}^m \lambda_i^h \nabla h_i(x) \\
&+ \sum_{i=1}^q \nabla G_i(x) \left(\lambda_i^G + \lambda_i^\Phi (F_{1i}(x;\hat{t}) - F_{2i}(x;\hat{t}) t\theta_r'(G_i(x))) - \nu_i^{F_1} t\theta_r'(G_i(x)) - \nu_i^{F_2} + \mu_i^{F_1} t\theta_r'(G_i(x)) + \mu_i^{F_2} \right) \\
&+ \sum_{i=1}^q \nabla H_i(x) \left(\lambda_i^H + \lambda_i^\Phi (F_{2i}(x;\hat{t}) - F_{1i}(x;\hat{t}) t\theta_r'(H_i(x))) - \nu_i^{F_2} t\theta_r'(H_i(x)) - \nu_i^{F_1} + \mu_i^{F_2} t\theta_r'(H_i(x)) + \mu_i^{F_1} \right),
\end{aligned}$$

with $\text{supp}(\lambda^g) \subset \mathcal{I}_g(x)$, $\text{supp}(\lambda^G) \subset \mathcal{I}_G(x;\hat{t})$, $\text{supp}(\lambda^H) \subset \mathcal{I}_H(x;\hat{t})$, $\text{supp}(\lambda^\Phi) \subset \mathcal{I}^{+0}_{GH}(x;\hat{t}) \cup \mathcal{I}^{0+}_{GH}(x;\hat{t})$, $\text{supp}(\nu^{F_1}) \subset I$, $\text{supp}(\nu^{F_2}) \subset I$ and $\text{supp}(\mu^{F_1}) \subset I^c$, $\text{supp}(\mu^{F_2}) \subset I^c$ where $I \cup I^c = \mathcal{I}^{00}_{GH}(x;\hat{t})$ and $I \cap I^c = \emptyset$.

Using the assumption of linear independence of the gradients gives that the solution of the equation above satisfy the following system of equations

$$\begin{aligned}
&\lambda^g = 0, \lambda^h = 0, \lambda^G = 0, \lambda^H = 0\\
&-\lambda^\Phi_i F_{2i}(x;\hat{t}) t\theta_r'(G_i(x)) = 0 \text{ and } \lambda^\Phi_i F_{2i}(x;\hat{t}) = 0 \ \forall i \in \mathcal{I}^{0+}_{GH}(x;\hat{t})\\
&\lambda^\Phi_i F_{1i}(x;\hat{t}) = 0 \text{ and } -\lambda^\Phi_i F_{1i}(x;\hat{t}) t\theta_r'(H_i(x)) = 0 \ \forall i \in \mathcal{I}^{+0}_{GH}(x;\hat{t})\\
&-\nu_i^{F_1} t\theta_r'(G_i(x)) - \nu_i^{F_2} = 0 \text{ and } -\nu_i^{F_2} t\theta_r'(H_i(x)) - \nu_i^{F_1} = 0 \ \forall i \in I\\
&\mu_i^{F_1} t\theta_r'(G_i(x)) + \mu_i^{F_2} = 0 \text{ and } \mu_i^{F_2} t\theta_r'(H_i(x)) + \mu_i^{F_1} = 0 \ \forall i \in I^c
\end{aligned}$$

From the second and third equations it follows that $\lambda^\Phi_i = 0$. The second last equation for $i \in I$ gives

$$\nu_i^{F_2} = -\nu_i^{F_1} t\theta_r'(G_i(x)).$$

This implies that $\nu_i^{F_1} = \nu_i^{F_2} = 0$ by non-decreasing hypothesis on θ and non-negativity of $\nu_i^{F_1}$ and $\nu_i^{F_2}$.

We proceed in the exact same way with the last equation to get $\mu_i^{F_1} = \mu_i^{F_2} = 0$.

This completes the proof that the trivial solution is the only solution to our first equation and so the results follows. □

Now we move to the theorem stating the constraint qualifications satisfied by the butterfly relaxations.

Theorem 4.2. *Let $x^* \in \mathcal{X}^B_{\hat{t}}$ such that MPCC-LICQ holds at x^*. Then, there exists $\bar{t} > 0$ and a neighbourhood $U(x^*)$ of x^* such that the following holds for all $t \in (0, \bar{t}]$. If $x \in U(x^*) \cap \mathcal{X}^B_{\hat{t}}$, then standard GCQ for $R^B_{\hat{t}}$ holds in x.*

Proof. First we note that it always holds that $\mathscr{L}^\circ_{\mathcal{X}^B_{\hat{t}}}(x) \subseteq \mathcal{T}^\circ_{\mathcal{X}^B_{\hat{t}}}(x)$. So, it sufficient to show the reverse inclusion.

The linearized cone of $R^B_{\hat{t}}$ is given by

$$\begin{aligned}
\mathscr{L}_{\mathcal{X}^B_{\hat{t}}}(x) = \{d \in \mathbb{R}^n \mid &\nabla g_i(x)^T d \le 0, \ i \in \mathcal{I}_g(x), \ \nabla h_i(x)^T d = 0, \ i = 1, \dots, m\\
&\nabla G_i(x)^T d \ge 0, \ i \in \mathcal{I}_G(x), \ \nabla H_i(x)^T d \ge 0, \ i \in \mathcal{I}_H(x)\\
&\nabla \Phi^B_i(G(x), H(x); \hat{t})(x)^T d \le 0, \ i \in \mathcal{I}^{0+}_{GH}(x;\hat{t}) \cup \mathcal{I}^{+0}_{GH}(x;\hat{t})\}.
\end{aligned}$$

Let us compute the polar of tangent cone. Consider the non linear program $(\text{NLP}_{t,I}(x))$ with $I \in \mathcal{P}(\mathcal{I}^{00}_{GH}(x;\hat{t}))$. By construction of $U(x^*)$ and $\bar{t}$, the gradients $\{\nabla g_i(x^*)\ (i \in \mathcal{I}_g(x^*)), \nabla h_i(x^*)\ (i = 1, \dots, m), \nabla G_i(x^*)\ (i \in \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*), \nabla H_i(x^*)\ (i \in \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*))\}$ remain linearly independent for all $x \in U(x^*)$ by continuity of the gradients in a neighbourhood and

$$\begin{aligned}
&\mathcal{I}_g(x) \subseteq \mathcal{I}_g(x^*)\\
&\mathcal{I}_G(x) \subseteq \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*)\\
&\mathcal{I}_H(x) \subseteq \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*)\\
&\mathcal{I}^{00}_{GH}(x;\hat{t}) \cup \mathcal{I}^{+0}_{GH}(x;\hat{t}) \subseteq \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*)\\
&\mathcal{I}^{00}_{GH}(x;\hat{t}) \cup \mathcal{I}^{0+}_{GH}(x;\hat{t}) \subseteq \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*).
\end{aligned}$$

Therefore, we can apply Lemma 4.2 that gives that MFCQ holds for $(\text{NLP}_{t,I}(x))$ at x. Furthermore, by Lemma 4.1 and since MFCQ in particular implies Abadie CQ it follows

$$\mathcal{T}_{\mathcal{X}^B_{\hat{t}}}(x) = \cup_{I \subseteq \mathcal{I}^{00}_{GH}(x;\hat{t})} \mathcal{T}_{NLP(\hat{t},I)}(x) = \cup_{I \subseteq \mathcal{I}^{00}_{GH}(x;\hat{t})} \mathscr{L}_{NLP(\hat{t},I)}(x).$$

By [[11], Theorem 3.1.9], passing to the polar yields

$$\mathcal{T}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x) = \cap_{I \subseteq \mathcal{I}^{00}_{GH}(x;\hat{t})} \mathscr{L}^{\circ}_{NLP(\hat{t},I)}(x),$$

and by [[11], Theorem 3.2.2]

$$\begin{aligned}\mathscr{L}^{\circ}_{NLP(\hat{t},I)}(x) = \{v \in \mathbb{R}^n \mid v = &\sum_{i\in\mathcal{I}_g(x)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^{m} \lambda^h_i \nabla h_i(x) - \sum_{i\in\mathcal{I}_G(x)} \lambda^G_i \nabla G_i(x) - \sum_{i\in\mathcal{I}_H(x)} \lambda^H_i \nabla H_i(x) \\ &+ \sum_{i\in\mathcal{I}^{+0}_{GH}(x;\hat{t})\cup\mathcal{I}^{0+}_{GH}(x;\hat{t})} \lambda^{\Phi}_i \nabla \Phi^B_i(G(x),H(x);\hat{t}) - \sum_{i\in I} \lambda^G_i \nabla G_i(x) + \sum_{i\in I^c} \lambda^G_i \nabla G_i(x) \\ &- \sum_{i\in I} \lambda^H_i \nabla H_i(x) + \sum_{i\in I^c} \lambda^H_i \nabla H_i(x) : \ \lambda^g, \lambda^G, \lambda^H, \lambda^{\Phi} \geq 0\}\end{aligned}$$

Taking $v \in \mathcal{T}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x)$ implies $v \in \mathscr{L}^{\circ}_{NLP(\hat{t},I)}(x)$ for all $I \subseteq \mathcal{I}^{00}_{GH}(x;\hat{t})$. If we fix such I, then there exists some multipliers λ^h and $\lambda^g, \lambda^G, \lambda^H, \lambda^{\Phi} \geq 0$ so that

$$\begin{aligned}v = &\sum_{i\in\mathcal{I}_g(x)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^{m} \lambda^h_i \nabla h_i(x) - \sum_{i\in\mathcal{I}_G(x)} \lambda^G_i \nabla G_i(x) - \sum_{i\in\mathcal{I}_H(x)} \lambda^H_i \nabla H_i(x) + \sum_{i\in\mathcal{I}^{+0}_{GH}(x;\hat{t})\cup\mathcal{I}^{0+}_{GH}(x;\hat{t})} \lambda^{\Phi}_i \nabla \Phi^B_i(G(x),H(x);\hat{t}) \\ &- \sum_{i\in I} \lambda^G_i \nabla G_i(x) + \sum_{i\in I^c} \lambda^G_i \nabla G_i(x) - \sum_{i\in I} \lambda^H_i \nabla H_i(x) + \sum_{i\in I^c} \lambda^H_i \nabla H_i(x).\end{aligned}$$

Now, it also holds that $v \in \mathscr{L}^{\circ}_{NLP(\hat{t},I^c)}(x)$ and so there exists some multipliers λ^h and $\lambda^g, \lambda^G, \lambda^H, \lambda^{\Phi} \geq 0$ such that

$$\begin{aligned}v = &\sum_{i\in\mathcal{I}_g(x)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^{m} \lambda^h_i \nabla h_i(x) - \sum_{i\in\mathcal{I}_G(x)} \lambda^G_i \nabla G_i(x) - \sum_{i\in\mathcal{I}_H(x)} \lambda^H_i \nabla H_i(x) + \sum_{i\in\mathcal{I}^{+0}_{GH}(x;\hat{t})\cup\mathcal{I}^{0+}_{GH}(x;\hat{t})} \lambda^{\Phi}_i \nabla \Phi^B_i(G(x),H(x);\hat{t}) \\ &+ \sum_{i\in I} \lambda^G_i \nabla G_i(x) - \sum_{i\in I^c} \lambda^G_i \nabla G_i(x) + \sum_{i\in I} \lambda^H_i \nabla H_i(x) - \sum_{i\in I^c} \lambda^H_i \nabla H_i(x).\end{aligned}$$

By construction of $\bar{t}$ and $U(x^*)$ the gradients involved here are linearly independent and so the multipliers in both previous equations must be equal. Thus, the multipliers λ^G_i and λ^H_i with indices i in $I \cup I^c$ vanish. Therefore, $v \in \mathscr{L}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x)$ and as v has been chosen arbitrarily then $\mathcal{T}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x) \subseteq \mathscr{L}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x)$.
The result follows since it always holds that $\mathscr{L}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x) \subseteq \mathcal{T}^{\circ}_{\mathcal{X}^B_{\hat{t}}}(x)$.

□

The following example shows that this results is sharp since Abadie CQ does not hold on Example 4.7.

Example 4.7.

$$\min_{(x_1,x_2)\in\mathbb{R}^2} f(x) \ s.t. \ 0 \leq x_1 \perp x_2 \geq 0.$$

At $x^* = (0,0)^T$ it holds that $\nabla\Phi^B(G(x),H(x);\hat{t})(x^*) = (0,0)^T$ and so $\mathscr{L}_{\mathcal{X}^B_{\hat{t}}}(x^*) = \mathbb{R}^2$, which is obviously different from the tangent cone at x^*.

Regarding the butterfly relaxation with $t\theta'(0) = r$ an improved regularity result holds.

Theorem 4.3. *Let $x^* \in \mathcal{X}^B_{t\theta'(0)=r}$ such that MPCC-LICQ holds at x^*. Then, there exists $\bar{t} > 0$ and a neighbourhood $U(x^*)$ of x^* such that the following holds for all $t \in (0,\bar{t}]$. If $x \in U(x^*) \cap \mathcal{X}^B_{t\theta'(0)=r}$, then standard ACQ for $R^B_{t\theta'(0)=r}$ holds in x.*

Proof. The proof follows exactly the ones of Theorem 4.2, until we compute the tangent cone of $\mathcal{X}^B_{t\theta'(0)=r}$ as

$$\mathcal{T}_{\mathcal{X}^B_{t\theta'(0)=r}}(x) = \cup_{I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})}\mathcal{T}_{NLP(t,I)}(x) = \cup_{I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})}\mathscr{L}_{NLP(t,I)}(x),$$

since in particular ACQ holds in x for $(\mathrm{NLP}_{t,I}(x))$. A simple computation gives the linearized tangent cone for $(\mathrm{NLP}_{t,I}(x))$

$$\begin{aligned}\mathscr{L}_{NLP_{t,I}}(x) = \{d\in\mathbb{R}^n \mid\ & \nabla g_i(x)^T d\le 0,\ i\in\mathcal{I}_g(x),\ \nabla h_i(x)^T d = 0,\ i=1,\dots,m\\ & \nabla G_i(x)^T d\ge 0,\ i\in\mathcal{I}_G(x),\ \nabla H_i(x)^T d\ge 0,\ i\in\mathcal{I}_H(x)\\ & \nabla F_1(x;r,t)_i(x;t)^T d\le 0,\ \nabla F_2(x;t,r)_i(x;t)\ge 0,\ i\in I\\ & \nabla F_1(x;r,t)_i(x;t)^T d\ge 0,\ \nabla F_2(x;t,r)_i(x;t)\le 0,\ i\in I^c\\ & \nabla\Phi^B_i(G(x),H(x);\hat{t})^T d\le 0,\ i\in\mathcal{I}^{0+}_{GH}(x;\hat{t})\cup\mathcal{I}^{+0}_{GH}(x;\hat{t})\}.\end{aligned}$$

Moreover one has for all $I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})$,

$$\begin{aligned}\nabla F_1(x;r,t)_i(x;t) &= \nabla H_i(x) - t\theta'_r(G_i(x))\nabla G_i(x)\\ &= \nabla H_i(x) - \nabla G_i(x) = -\nabla F_2(x;t,r)_i(x;t),\end{aligned}$$

since $t\theta'(0)=r$ and for all $i\in\mathcal{I}^{00}_{GH}(x;\hat{t})$ it follows that $G_i(x)=H_i(x)=0$.

It is to be noted that

$$\cup_{I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})}\{d\in\mathbb{R}^n \mid \nabla H_i(x)^T d\le\nabla G_i(x),\ i\in I\ ;\ \nabla H_i(x)^T d\ge\nabla G_i(x),\ i\in I^c\} = \mathbb{R}^n.$$

Therefore, it holds that

$$\begin{aligned}\cup_{I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})}\mathscr{L}_{NLP(t,I)}(x) = \{d\in\mathbb{R}^n \mid\ & \nabla g_i(x)^T d\le 0,\ i\in\mathcal{I}_g(x),\ \nabla h_i(x)^T d = 0,\ i=1,\dots,m\\ & \nabla G_i(x)^T d\ge 0,\ i\in\mathcal{I}_G(x),\ \nabla H_i(x)^T d\ge 0,\ i\in\mathcal{I}_H(x)\\ & \nabla\Phi^B_i(G(x),H(x);\hat{t})^T d\le 0,\ i\in\mathcal{I}^{0+}_{GH}(x;\hat{t})\cup\mathcal{I}^{+0}_{GH}(x;\hat{t})\}.\end{aligned}$$

Thus, $\cup_{I\subseteq\mathcal{I}^{00}_{GH}(x;\hat{t})}\mathscr{L}_{NLP(t,I)}(x) = \mathscr{L}_{\mathcal{X}^B_{t\theta'(0)=r}}(x)$ and the result follows. □

This result is sharp since very weak constraint qualification like quasinormality and CCP does not hold in example 4.7.

Example 4.8. *We show that CRSC and CCP do not hold in $x^*=(0,0)^T$ for the relaxation $R^B_{t\theta'(0)=r}$.*

It holds that $\nabla\Phi^B_{t\theta'(0)=r}(x^*) = (0,0)^T$, therefore $\nabla\Phi^B_{t\theta'(0)=r}(x^*)\in -\mathscr{L}(x^*)^\circ$. However, for any x in a small neighbourhood around x^* the gradient $\nabla\Phi^B_{t\theta'(0)=r}(x^*)\neq(0,0)^T$. So, the rank is not constant in this neighbourhood and CRSC does not hold in x^*.

In order to verify that CCP does not hold in x^*, we check that there exists an AKKT sequence that does not converges to a KKT point. Given $\{x^k\}$, $\{\lambda^{G,k}\},\{\lambda^{H,k}\},\{\lambda^{\Phi,k}\}$ such that

$$\begin{aligned}&\lim_{k\to\infty}\nabla f(x^k) + \lambda^{\Phi,k}\nabla\Phi^B_{(t\theta'(0)=r)}(x^k) - \lambda^{G,k}\nabla G(x^k) - \lambda^{H,k}\nabla H(x^k) = 0\\ &\lim_{k\to\infty}\min(\lambda^{\Phi,k}, -\Phi^B_{(t\theta'(0)=r)}(x^k)) = 0\\ &\lim_{k\to\infty}\min(\lambda^{G,k}, G(x^k)) = 0\\ &\lim_{k\to\infty}\min(\lambda^{H,k}, H(x^k)) = 0.\end{aligned}$$

We can find a sequence x^k such that $t_k\theta'_{r_k}(H(x^k))\to 0$, $F_1(x^k;\hat{t}_k)=0$, $F_2(x^k;\hat{t}_k)\ge 0$ and $\lambda^k = \frac{1}{kF_2(x^k;\hat{t}_k)}$, $\lambda^{G,k}\to\frac{1}{k}$, $\lambda^{H,k}\to 0$ so that CCP does not hold at x^*.

The following example shows that we can not have a similar result with MPCC-MFCQ instead of MPCC-LICQ for Theorem 4.3.

Example 4.9. *Consider*

$$0 \leq x_1 + x_2^2 \perp x_1 \geq 0.$$

MPCC-MFCQ holds at x^*, since the gradients are linearly dependent but only with coefficients $\lambda^G = -\lambda^H$ that does not respect the condition given in Definition 2.10.

Now, taking a sequence of stationary point such that $x^k \to x^* = (0,0)^T$ and

$$F_2(x^k; \hat{t}_k) = 0, -t_k \theta'_{r_k}(H(x^k)) \to -1.$$

Since $\nabla G(x^*) = \nabla H(x^*)$ it holds that $\nabla F_2(x;t,r)(x^*;0) = (0 \ 0)^T$ and so MFCQ does not hold for $(\mathrm{NLP}_{t,I}(x))$.

Both Theorem 4.2 and Theorem 4.3 are slightly disappointing since MPCC-LICQ is a quite strong assumption. Fortunately, the following result guarantees that the difficulties are only localized in indices i of x^* that belongs to $\mathcal{I}^{00}(x^*)$.

Theorem 4.4. *Let $x^* \in \mathcal{X}^B_{\hat{t}}$ be such that MPCC-LICQ holds at x^*. Then, there exists $\bar{t} > 0$ and a neighbourhood $U(x^*)$ of x^* such that the following holds for all $t \in (0, \bar{t}]$. If $x \in U(x^*) \cap \mathcal{X}^B_{\hat{t}}$ and $\mathcal{I}^{00}_{GH}(x;\hat{t}) = \emptyset$, then standard LICQ for $R^B_{\hat{t}}$ holds in x.*

Proof. Following the same path than Lemma 4.2, the gradient of the Lagrangian for $R^B_{\hat{t}}$ at $x \in U(x^*)$ for $\hat{t}$ sufficiently small gives

$$\begin{aligned} 0 = &\sum_{i \in \mathcal{I}_g(x)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^{m} \lambda^h_i \nabla h_i(x) + \sum_{i \in \mathcal{I}_G(x)} \nabla G_i(x) \lambda^G_i + \sum_{i \in \mathcal{I}_H(x)} \nabla H_i(x) \lambda^H_i \\ &+ \sum_{i \in \mathcal{I}^{+0}_{GH}(x;\hat{t})} \nabla G_i(x) \left(\lambda^\Phi_i (F_{1i}(x;\hat{t}) - F_{2i}(x;\hat{t}) t \theta'_r(G_i(x))) \right) + \sum_{i \in \mathcal{I}^{0+}_{GH}(x;\hat{t})} \nabla H_i(x) \left(\lambda^\Phi_i (F_{2i}(x;\hat{t}) - F_{1i}(x;\hat{t}) t \theta'_r(H_i(x))) \right). \end{aligned}$$

Using the assumption of linear independence of the gradients and

$$\begin{aligned} &\mathcal{I}_g(x) \subseteq \mathcal{I}_g(x^*) \\ &\mathcal{I}_G(x) \subseteq \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*) \\ &\mathcal{I}_H(x) \subseteq \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*) \\ &\mathcal{I}^{+0}_{GH}(x;\hat{t}) \subseteq \mathcal{I}^{00}(x^*) \cup \mathcal{I}^{0+}(x^*) \\ &\mathcal{I}^{0+}_{GH}(x;\hat{t}) \subseteq \mathcal{I}^{+0}(x^*) \cup \mathcal{I}^{00}(x^*) \end{aligned}$$

gives that the solution of the equation above satisfy the following system of equations

$$\begin{aligned} &\lambda^g = 0, \lambda^h = 0, \lambda^G = 0, \lambda^H = 0 \\ &-\lambda^\Phi_i F_{2i}(x;\hat{t}) t \theta'_r(G_i(x)) = 0 \text{ and } \lambda^\Phi_i F_{2i}(x;\hat{t}) = 0 \ \forall i \in \mathcal{I}^{0+}_{GH}(x;\hat{t}) \\ &\lambda^\Phi_i F_{1i}(x;\hat{t}) = 0 \text{ and } -\lambda^\Phi_i F_{1i}(x;\hat{t}) t \theta'_r(H_i(x)) = 0 \ \forall i \in \mathcal{I}^{+0}_{GH}(x;\hat{t}) \end{aligned}$$

From the second and third equations it follows that $\lambda^\Phi_i = 0$. So, the only solution is the trivial solution. Thus, the result follows. □

4.3 Convergence of the ϵ-stationary points

Non-linear programming algorithms usually compute sequences of approximate stationary points or ϵ-stationary points. This approach that has become an active subject recently alter significantly the convergence analysis of relaxation methods as stated in [46, 49, 50] and [63].

Previous results in the literature in [50] provides convergence to C-stationary point for the relaxation (SS) and the one from Lin & Fukushima, [54], at the limit point and under hypothesis on the sequence ϵ_k, respectively $\epsilon_k = O(t_k)$ and $\epsilon_k = o(t_k^2)$. Furthermore, they provide a counter-example with sequences converging to a weak-stationary point if this conditions does not hold. Although in [50], they prove that relaxation (SU), (KDB) and (KS) converge only to a weak stationary point and require more hypothesis on the sequences ϵ_k and x_k to improve to a C- or an M-stationary limit point. In the following theorem we prove that the situation is similar with the new butterfly relaxation method.

Lemma 4.3. *Given $\{\hat{t}_k\}$ a sequence of parameter satisfying* (6) *and $\{\epsilon_k\}$ a sequence of non-negative parameter such that both sequences decrease to zero as $k \in \mathbb{N}$ goes to infinity. Let $\{x^k, \lambda^k\}_k$ be a sequence of ϵ_k-stationary points of $(R^B_{\hat{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$. Let $\{\eta^{G,k}\}, \{\eta^{H,k}\}$ be two sequences such that*

$$\begin{aligned}
\eta_i^{G,k} &:= \begin{cases} \lambda_i^{G,k} + \lambda_i^{\Phi,k}\left(t_k\theta'_{r_k}(G_i(x^k))F_{2i}(x^k;\hat{t}_k) - F_{1i}(x^k;\hat{t}_k)\right) & \text{if } F_{1i}(x^k;\hat{t}_k) + F_{2i}(x^k;\hat{t}_k) \geq 0 \\ \lambda_i^{G,k} + \lambda_i^{\Phi,k}\left(F_{2i}(x^k;\hat{t}_k) - t_k\theta'_{r_k}(G_i(x^k))F_{1i}(x^k;\hat{t}_k)\right) & \text{if } F_{1i}(x^k;\hat{t}_k) + F_{2i}(x^k;\hat{t}_k) < 0 \end{cases} \\
\eta_i^{H,k} &:= \begin{cases} \lambda_i^{H,k} + \lambda_i^{\Phi,k}\left(t_k\theta'_{r_k}(H_i(x^k))F_{1i}(x^k;\hat{t}_k) - F_{2i}(x^k;\hat{t}_k)\right) & \text{if } F_{1i}(x^k;\hat{t}_k) + F_{2i}(x^k;\hat{t}_k) \geq 0 \\ \lambda_i^{H,k} + \lambda_i^{\Phi,k}\left(F_{1i}(x^k;\hat{t}_k) - t_k\theta'_{r_k}(H_i(x^k))F_{2i}(x^k;\hat{t}_k)\right) & \text{if } F_{1i}(x^k;\hat{t}_k) + F_{2i}(x^k;\hat{t}_k) < 0 \end{cases}
\end{aligned}, \tag{10}$$

for $i \in \{1, \ldots, q\}$. Assume that the sequence of multipliers $\{\lambda^{h,k}, \lambda^{g,k}, \eta^{G,k}, \eta^{H,k}\}$ is bounded. Then, x^* is a weak-stationary point of (MPCC).

About the choice of the sequence ϵ_k it is of interest to see that $\epsilon_k = o(r_k)$ implies that $\epsilon_k = o(\bar{r}_k)$ by definition of the latter.

Proof. By definition, since x^k is an ϵ_k stationary point for $R^B_{\hat{t}}$ it holds for all $k \in \mathbb{N}$

$$\begin{aligned}
\Bigg\| \nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^{g,k}\nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k}\nabla h_i(x^k) - \sum_{i=1}^{q} \lambda_i^{G,k}\nabla G_i(x^k) \\
- \sum_{i=1}^{q} \lambda_i^{H,k}\nabla H_i(x^k) + \sum_{i=1}^{q} \lambda_i^{\Phi,k}\nabla\Phi_i^B(G(x), H(x); \hat{t})(x^k) \Bigg\|_\infty \leq \epsilon_k
\end{aligned}$$

with

$$\begin{aligned}
&|h_i(x^k)| \leq \epsilon_k, \ \forall i \in \{1, \ldots, m\} \\
&g_i(x^k) \leq \epsilon_k, \ \lambda_i^{g,k} \geq 0, \ \left|\lambda_i^{g,k} g_i(x^k)\right| \leq \epsilon_k \ \forall i \in \{1, \ldots, p\} \\
&G_i(x^k) + \bar{r}_k \geq -\epsilon_k, \ \lambda_i^{G,k} \geq 0, \ \left|\lambda_i^{G,k}(G_i(x^k) + \bar{r}_k)\right| \leq \epsilon_k \ \forall i \in \{1, \ldots, q\} \\
&H_i(x^k) + \bar{r}_k \geq -\epsilon_k, \ \lambda_i^{H,k} \geq 0, \ \left|\lambda_i^{H,k}(H_i(x^k) + \bar{r}_k)\right| \leq \epsilon_k \ \forall i \in \{1, \ldots, q\} \\
&\Phi_i^B(G(x), H(x); \hat{t})(x^k) \leq \epsilon_k, \ \lambda_i^{\Phi,k} \geq 0, \ \left|\lambda_i^{\Phi,k}\Phi_i^B(G(x), H(x); \hat{t})(x^k)\right| \leq \epsilon_k \ \forall i \in \{1, \ldots, q\}.
\end{aligned}$$

The representation of $\Phi_i^B(G(x), H(x); \hat{t})(x^k)$ immediately gives $\nabla\Phi_i^B(G(x), H(x); \hat{t})(x^k) = 0, \ \forall i \in \mathcal{I}_{GH}^{00}(x^k; \hat{t}_k)$ for all $k \in \mathbb{N}$. Thus, we can rewrite the equation above as

$$\begin{aligned}
\Bigg\| \nabla f(x^k) + \sum_{i=1}^{p} \lambda_i^{g,k}\nabla g_i(x^k) + \sum_{i=1}^{m} \lambda_i^{h,k}\nabla h_i(x^k) - \sum_{i=1}^{q} \eta_i^{G,k}\nabla G_i(x^k) \\
- \sum_{i=1}^{q} \eta_i^{H,k}\nabla H_i(x^k) \Bigg\|_\infty \leq \epsilon_k.
\end{aligned}$$

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

$$\{\lambda^{h,k}, \lambda^{g,k}, \eta^{G,k}, \eta^{H,k}\} \to (\lambda^h, \lambda^g, \eta^G, \eta^H).$$

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

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

We prove that $(x^*, \lambda^h, \lambda^g, \eta^G, \eta^H)$ is a weak-stationary point. Obviously, since $\epsilon_k \downarrow 0$ it follows that $x^* \in \mathcal{Z}$, $\nabla_x \mathcal{L}^1_{MPCC}(x^*, \lambda^h, \lambda^g, \eta^G, \eta^H) = 0$ by previous inequality and that $\lambda^g_i = 0$ for $i \notin \mathcal{I}_g(x^*)$. It remains to show that for indices $i \in \mathcal{I}^{+0}(x^*)$, $\eta^G_i = 0$. 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 ϵ_k-stationarity it holds for all k that

$$|\lambda^{G,k}_i (G_i(x^k) + \bar{t_k})| \leq \epsilon_k.$$

Therefore, $\lambda^{G,k}_i \to_{k\to\infty} 0$ since $\epsilon_k \downarrow 0$ and $G_i(x^k) \to G_i(x^*) > 0$.
Now, there is two possible cases either $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$ either $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) < 0$.

Consider the case $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$ and denote

$$\begin{aligned}
\alpha^H(x^k_i; t_k) &:= -t_k \theta'_{r_k}(G_i(x^k)) F_{2i}(x^k; \hat{t}_k) + F_{1i}(x^k; \hat{t}_k)\\
\alpha^G(x^k_i; t_k) &:= -t_k \theta'_{r_k}(H_i(x^k)) F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k).
\end{aligned}$$

It remains to prove that $\alpha^H(x^k_i; t_k)\lambda^{\Phi,k}_i \to_{k\to\infty} 0$. Assume by contradiction that

$$\lim_{k\to\infty} \alpha^H(x^k_i; t_k)\lambda^{\Phi,k}_i = C < 0, \tag{11}$$

which is necessary a finite value by boundedness hypothesis of the sequence of multiplier. Obviously the sequence $\{\lambda^{\Phi,k}\}$ must be unbounded otherwise (11) does not hold.

Additionally, $\lim_{k\to\infty} F_{1i}(x^k; \hat{t}_k)\lambda^{\Phi,k}_i = 0$ since $\left|\lambda^{\Phi,k}_i \Phi^B_i(G(x), H(x); \hat{t})(x^k)\right| \leq \epsilon_k$. So, by (5) we have that $\lim\limits_{k\to\infty} \alpha^G_i(x^k; t_k) = G_i(x^*) > 0$ and therefore $\lim\limits_{k\to\infty} \alpha^G(x^k; t_k)\lambda^{\Phi,k}_i = \lim\limits_{k\to\infty} G_i(x^k)\lambda^{\Phi,k}_i = \infty$. Boundedness assumption in the statement of the theorem implies that η^H_i is bounded and so

$$\lim_{k\to\infty} |\lambda^{H,k}_i - \alpha^G(x^k; t_k)\lambda^{\Phi,k}_i| < \infty.$$

The complementarity conditions on $H_i(x^k) \geq -r\bar{}_k$ gives that necessary $H_i(x^k) \sim -r\bar{}_k$ otherwise $\lambda^{H,k}_i$ would be unbounded.

However, this leads to a contradiction with $\lambda^{\Phi,k} \to \infty$, since $\lambda^{\Phi,k}_i F_1(x; r, t)(x^k) \to 0$ gives that $\lambda^{\Phi,k}_i r\bar{}_k \leq \epsilon_k$ and we assume in the statement of the theorem that $\epsilon_k = o(r_k)$. So in the case $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$ it holds that $\eta^{G,*}_i = 0$.

Let us consider the case $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) < 0$. As pointed out above it is true by (5) that $F_{1i}(x^k; \hat{t}_k) \to H_i(x^*)$ and $F_{2i}(x^k; \hat{t}_k) \to G_i(x^*)$. Therefore, for k sufficiently large this case never happen since we choose $i \in \mathcal{I}^{0+}(x^*)$.

This conclude the proof that $\eta^{G,*}_i = 0$. The case $i \in \mathcal{I}^{0+}(x^*)$ is completely similar by symmetry and gives that $\eta^{H,*}_i = 0$ for $i \in \mathcal{I}^{+0}(x^*)$. So, x^* is a weak-stationary point. □

Lemma 4.4. *Given $\{\hat{t}_k\}$ a sequence of parameter satisfying* (6) *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 $\epsilon_k = o(\max(G(x^k), H(x^k)))$ and $t_k = o(r_k)$. Let $\{x^k, \lambda^k\}_k$ be a sequence of ϵ_k-stationary points of $(R^B_{\hat{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$. Let $\{\eta^{G,k}\}, \{\eta^{H,k}\}$ be two sequences defined in* (10). *Assume that the sequence of multipliers $\{\lambda^{h,k}, \lambda^{g,k}, \eta^{G,k}, \eta^{H,k}\}$ is bounded. Then, x^* is an M-stationary point of* (MPCC).

Proof. First part of this proof that shows convergence of the sequence and weak-stationarity of x^* given by Lemma 4.3.
We now consider indices $i \in \mathcal{I}^{00}(x^*)$. Our aim here is to prove that x^* is an M-stationary point. Assume by contradiction that there exists a subsequence $K \subseteq \mathbb{N}$ and $i \in \mathcal{I}^{00}(x^*)$ such that $\eta_i^{G,k}\eta_i^{H,k} \to_K \eta_i^{G,*}\eta_i^{H,*} \neq 0$ and either $\eta_i^{G,k} < 0$ or $\eta_i^{H,k} < 0$.

Without loss of generality suppose that $\frac{\epsilon_k}{|G_i(x^k)|} \to_K 0$. Indeed, the case $\frac{\epsilon_k}{|H_i(x^k)|} \to_K 0$ is similar by symmetry and the case $\frac{\epsilon_k}{|H_i(x^k)|} \to_K > 0, \frac{\epsilon_k}{|G_i(x^k)|} \to_K > 0$ is not possible by definition of ϵ_k.

Let α and β be such that $\frac{G_i(x^k)}{t_k\theta_{r_k}(H_i(x^k))} \to \alpha$ and $\frac{H_i(x^k)}{t_k\theta_{r_k}(G_i(x^k))} \to \beta$. Obviously the sequence $\{\lambda^{\Phi,k}\}$ must be unbounded otherwise x^* is an S-stationarity by non-negativity of $\lambda^{G,k}$ and $\lambda^{H,k}$.

Consider the case with $\alpha \neq 1$ and $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$ then

$$\left|\lambda_i^{\Phi,k} F_{1i}(x^k; \hat{t}_k)\left(1 - \frac{t_k\theta_{r_k}(H_i(x^k)}{G_i(x^k)}\right)\right| \leq \left|\frac{\epsilon_k}{G_i(x^k)}\right|,$$

and so $\lambda_i^{\Phi,k} F_{1i}(x^k; \hat{t}_k) \to 0$. By hypothesis and by definition of ϵ_k we necessarily have $-G_i(x^k) \sim_K \bar{r}_k$ and $F_{2i}(x^k; \hat{t}_k) > 0$ otherwise $\lambda_i^{G,k} \to 0$ and then $\eta_i^{G,*}\eta_i^{H,*} = 0$. Therefore, it holds that $G_i(x^k) > t_k\theta_{r_k}(H_i(x^k)) \sim H_i(x^k)\frac{t_k}{r_k}\theta'(0)$ with $G_i(x^k) \leq 0$ and so, $H_i(x^k) = \Omega(r_k)$ since $\bar{r}_k \sim_K \frac{r_k^2}{t_k}$ when $t_k = o(r_k)$. In these conditions, $F_1(x; r, t)_i(x^k; \hat{t_k}) \sim_K r_k = \omega(\epsilon_k)$ and so $\lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k) \to 0$. However, this leads to a contradiction with $\lambda^{\Phi,k}$ unbounded and $\alpha \neq 1$.

In the case with $\alpha \neq 1$ and $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \leq 0$ then

$$\left|\lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k)\left(1 - \frac{t_k\theta_{r_k}(H_i(x^k)}{G_i(x^k)}\right)\right| \leq \left|\frac{2\epsilon_k}{G_i(x^k)}\right|,$$

and so $\lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k) \to 0$. Furthermore, it necessarily holds that $F_{1i}(x^k; \hat{t}_k) \leq 0$ otherwise $\eta_i^{G,*}\eta_i^{H,*} \geq 0$. We now consider two cases :

1. If $G_i(x^k) \geq 0$, then by $|\lambda_i^{G,k}| \leq \epsilon_k / G_i(x^k)$ it follows that $\lambda_i^{G,k} \to 0$ and so $\eta_i^{G,k} \to 0$.
2. If $G_i(x^k) \leq 0$, then in a similar way as in the case $\alpha \neq 1$ and $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$, it is necessary that $-G_i(x^k) \sim_K \bar{r}$ otherwise $\lambda_i^{G,k} \to 0$ and so by $F_{1i}(x^k; \hat{t}_k) \leq 0$ we have $H_i(x^k) = \Omega(r_k)$. Furthermore, we have $|H_i(x^k)| \leq \bar{r} + \epsilon$ and then $F_{1i}(x^k; \hat{t}_k) = \omega(r_k)$. The contradiction follows from the fact that $\lambda_i^{\Phi,k} F_{1i}(x^k; \hat{t}_k) \to 0$ since in these conditions $F_{1i}(x^k; \hat{t}_k) = \omega(\epsilon_k)$.

In both cases we get a contradiction.

Now, if $\alpha = 1$ then necessarily $\beta \neq 1$. Indeed, the case $\alpha = \beta = 1$ is not possible since it implies that

$$G_i(x^k) \sim t_k\theta_{r_k}(H_i(x^k)) \text{ and } H_i(x^k) \sim t_k\theta_{r_k}(G_i(x^k)),$$

which is in contradiction with $t_k = o(r_k)$. Considering $\beta \neq 1$ and $F_{1i}(x^k; \hat{t}_k) + F_{2i}(x^k; \hat{t}_k) \geq 0$ it follows

$$\left|\lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k)\left(1 - \frac{t_k\theta_{r_k}(G_i(x^k)}{H_i(x^k)}\right)\right| \leq \left|\frac{\epsilon_k}{H_i(x^k)}\right|.$$

$\left|\frac{\epsilon_k}{H_i(x^k)}\right| \to 0$ since by $\alpha = 1$ we have $|G_i(x^k)| < |H_i(x^k)$. It follows that $\lambda_i^{\Phi,k} F_{2i}(x^k; \hat{t}_k) \to 0$. Now, we consider two cases :

1. $H_i(x^k) \geq 0$: By $|\lambda_i^{H,k}(H_i(x^k) + \bar{r}_k)| \leq \epsilon_k$ and $\epsilon_k = o(\bar{r}_k)$, we have $\lambda_i^{H,k} \to 0$ and so $\eta_i^{H,k} \to 0$.

2. $H_i(x^k) \leq 0$: In order not to violate our hypothesis, $F_{1i}(x^k; \hat{t}_k)$ must be non-negative. That is $H_i(x^k) \geq t_k \theta_{r_k}(G_i(x^k))$ and so $|H_i(x^k)| \leq |G_i(x^k)|$, which is a contradiction with $\alpha = 1$.

In every cases we get a contradiction and so x^* is an M-stationary point. □

Theorem 4.5. *Given two sequences $\{t_k\}$ and $\{r_k\}$ satisfying (6) such that $\forall k \in \mathbb{N}, (t_k, r_k) \in \mathbb{R}^2_+$, both sequence decrease to zero as k goes to infinity. Let $\{x^k, \lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\Phi,k}\}$ be a sequence of points that are ϵ_k-stationary points of $R^B_{\hat{t}}(x^k)$ for all $k \in \mathbb{N}$ with $x^k \to x^*$ such that MPCC-CRSC holds at x^*. Furthermore assume that $t_k = o(r_k)$ $\forall k \in \mathbb{N}$ sufficiently large and that the sequence $\{\epsilon_k\}$ is such that $\epsilon_k = o(\max(G(x^k), H(x^k))$ and $\epsilon_k = o(r_k)$. Then, x^* is an M-stationary point.*

Proof. The proof is direct by Lemma 4.4 and Corollary 2.2 that ensures boundedness of the sequence (10) under MPCC-CRSC. □

In the weaker conditions of Lemma 4.3 boundedness of the sequence should be expected under MPCC-CRCQ in similar way as Proposition 4.1.

The following example from [48] shows that the butterfly relaxations with $t_k = o(r_k)$ may converge to an undesirable A-stationary point without the hypothesis that $\epsilon_k = o(\max(G(x^k), H(x^k))$.

Example 4.10.

$$\min_{(x_1,x_2)^T \in \mathbb{R}^2} x_2 - x_1 \quad s.t. \quad 0 \leq x_1 \perp x_2 \geq 0.$$

Let $t_k = r_k^2$ and choose any positive sequences $\{r_k\}$ and $\{\epsilon_k\}$ such that $r_k, \epsilon_k \to 0$. Consider the following ϵ-stationary sequence

$$x^k = (\epsilon_k, \frac{\epsilon_k}{2})^T, \ \lambda^{G,k} = 0, \ \lambda^{H,k} = 1 - \lambda^{\Phi,k}(r_k^2 \theta_{r_k}\left(\frac{\epsilon_k}{2}\right) F_1(x^k; \hat{t}_k)) - F_2(x^k; \hat{t}_k))$$

and

$$\lambda^{\Phi,k} = \frac{1}{r_k^2 \theta_{r_k}(\epsilon_k) F_2(x^k; \hat{t}_k)) - F_1(x^k; \hat{t}_k)}.$$

This sequence converge to $x^* = (0, 0)$, which is an A-stationary point.

We see on Figure 5 that the butterfly relaxations with ϵ feasibility is similar to relaxation (SS). Therefore, it is not surprising that we can only expect to converges to a C-stationary point without strong hypothesis. Those issues clearly deserves a specific studies that is left here for further research.

5 Numerical Results

In this section, we focus on the numerical implementation of the butterfly relaxation methods. Our aim is to compare these new methods with the existing ones in the literature and show some of their features. This comparison uses the collection of test problems MacMPEC [51]. This collection has been widely used in the literature to compare relaxation methods as in [36, 46, 68]. The test problems included in MacMPEC are extracted from the literature and real world applications.

We also present an example of an (MPCC) that illustrates the difficulties that may occur by dealing with ϵ-stationary points.

Finally, an adaptation of the butterfly relaxations to the mathematical program with vanishing constraints is presented.

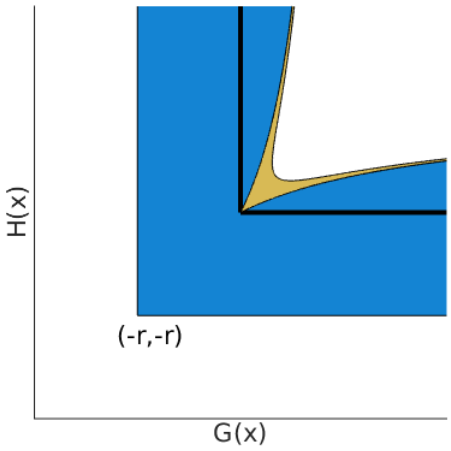

Figure 5: Feasible set of the butterfly relaxation with $2t = r$ and a constraint $\Phi^B(G(x), H(x); \hat{t}) \leq \epsilon$.

5.1 On the Implementation of the Butterfly Relaxations

As pointed out through this paper, the butterfly relaxations handle two parameters t and r. It can be practical to choose a relation between both parameters. Among the infinite possibilities of relationship between t and r, at least two are specific :

(i) $t = r$, since as stated in Theorem 4.3 this relaxation is more regular, but may converge to undesirable A-stationary points, Theorem 4.1;

(ii) $t = o(r)$, for instance $t = r^{3/2}$, which ensures convergence to M-stationary points as stated in Corollary 4.1.

Practical implementation could consider a slightly different model, by skipping the relaxation of the positivity constraint and adding a new parameter s in order to move the intersection of both wings in the point $(G(x), H(x)) = (s, s)$. This can be done by redefining $F_1(x; r, t)$ and $F_2(x; t, r)$ such that

$$F_{1i}(x; r, s, t) = (H_i(x) - s - t\theta_r(G_i(x) - s))$$
$$F_{2i}(x; r, s, t) = (G_i(x) - s - t\theta_r(H_i(x) - s)).$$

Even if we did not give any theoretical proof regarding this modified system, this modification does not alter the behaviour of the butterfly relaxations.

The numerical comparison of the butterfly relaxations with other existing methods considers the three schemes illustrated in Figure 6 :

1. $B_{(t=r)}$: $(s = 0, t = r)$
2. $B_{(t=r^{3/2})}$: $(s = 0, t = r^{3/2})$
3. $B_{(s=t,2t=r)}$: $(s = t, 2t = r)$

Each scheme is duplicated whether we omit or not the relaxation of the positivity constraints. The models with this relaxation are denoted $B^+_{(t=r)}$ with a $+$ as a superscript. So, we use 6 different butterfly schemes with very different properties.

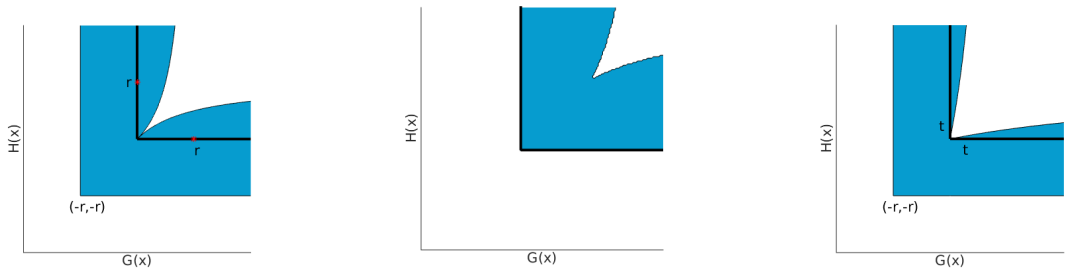

Figure 6: butterfly relaxations from left to right :$B_{(t=r)}$, $B_{(s=t,2t=r)}$ and $B_{(t=r^{3/2})}$.

5.2 Comparison of the Relaxation Methods

We provide in this section and in Algorithm 2 some more details on the implementation and the comparison between relaxation methods. It is to be noted that our aim is to compare the methods and so no attempt to optimize any method has been carried out. We use 101 test problems from MacMPEC, where are omitted the problems that exceed the limit of 300 variables or constraints and some problems with evaluation error of the objective function or the constraints. Algorithm 2 is coded in Matlab and uses the AMPL API.

$R_{\hat{t}_k}$ denotes the relaxed non-linear program associated with a generic relaxation, where except from the butterfly methods the parameter r_k does not play any role. At each step we compute x^{k+1} as a solution of $R_{\hat{t}_k}$ starting from x^k. Therefore, at each step the initial point is more likely to be infeasible for $R_{\hat{t}_k}$. The iterative process stops when t_k and r_k are smaller than some tolerance, denoted $p_{\min}$ which is set as 10^{-15} here, or when the solution x^{k+1} of $R_{\hat{t}_k}$ is considered an ϵ-solution of (MPCC). To consider x^{k+1} as an ϵ-solution with ϵ set as 10^{-7} we check three criteria :

a) Feasibility of the last relaxed non-linear program: $\nu_f(x) := \max(-g(x), |h(x)|, -\Phi(x))$,

b) Feasibility of the complementarity constraint: $\nu_{comp}(x) := \min(G(x), H(x))^2$,

(c) The complementarity between the Lagrange multipliers and the constraints of the last relaxed non-linear program :

$$\nu_c(x) := \max\left(\|g(x) \circ \lambda^g\|_\infty, \|h(x) \circ \lambda^h\|_\infty, \|G(x) \circ \lambda^G\|_\infty, \|H(x) \circ \lambda^H\|_\infty, \|\Phi^B(G(x), H(x); \hat{t})(x) \circ \lambda^\Phi\|_\infty\right)$$

Obviously, it is hard to ask a tighter condition on the complementarity constraint since the feasibility only guarantees that the product component-wise with less than ϵ. Using these criteria we define a measure of optimality

$$min\_local(x) := \max\left(\nu_f(x), \nu_{comp}(x), \nu_c(x)\right).$$

A fourth criterion could be the dual feasibility, that is the norm of the Lagrangian. However, solvers like SNOPT or MINOS do not use this criterion as the stopping criterion. One reason among other to discard such a criteria could be numerical issues implied by the degeneracy in the KKT conditions.

In the case of an infeasible or unbounded sub-problem $R_{\hat{t}_k}$, the algorithm stops and return a certificate.

Data:
starting vector x^0; initial relaxation parameter t_0; update parameter $(\sigma_t, \sigma_r) \in (0,1)^2$ and p_{min} the minimum parameter value, ϵ the precision tolerance ;
1 **Begin** ;
2 Set $k := 0$;
3 **while** $\max(t_k, r_k) > p_{\min}$ ***and*** $min\_local(x) > \epsilon$ **do**
4 | x^{k+1} solution of R_{t_k,r_k} with x^k initial point;
5 | $(t_{k+1}, r_{k+1}) := (t_k\sigma_t, r_k\sigma_r)$;
6 **return**: f_{opt} the optimal value at the solution x_{opt} or a decision of infeasibility or unboundedness.

Algorithm 2: Basic Relaxation methods for (MPCC), with a relaxed non-linear program $R_{\hat{t}_k}$.

Step 4 in Algorithm 2 is performed using three different solvers accessible through AMPL, [27], that are SNOPT 7.2-8 [29], MINOS 5.51 [59] and IPOPT 3.12.4 [70] with their default parameters. Previous similar comparison in the literature only consider SNOPT to solve the sub-problems.

We compare the butterfly schemes with the relaxations (SS) and (KS) that we respectively denote SS and KS. Moreover, we also take into account results of the non-linear programming solver without specific MPCC tuning and denote it NL.

In order to compare the various relaxation methods we need to have a coherent use of the parameters. In a similar way as in [67] we consider the value of the "intersection between G and H", which is (t, t) for KDB, KS and Butterfly , $(\sqrt{t}, \sqrt{t})$ for SS and $\frac{2\pi}{\pi-2}(t, t)$ for SU. Then, we run a sensitivity analysis on several values of the parameters $T \in \{100, 25, 10, 5, 1, 0.5, 0.05\}$ and $S \in \{0.1, 0.075, 0.05, 0.025, 0.01\}$, which corresponds to t_0 and σ_t as described in Table 1.

Relaxation	NL	SS	KS	Butterfly
t_0	none	T^2	T	T
σ_t	none	S^2	S	S

Table 1: Parameter links among the methods

SS	KS	Butterfly
$(\sqrt{t}, \sqrt{t})$	(t, t)	(t, t)

Table 2: Link between the parameters of relaxation methods.

In [39], the authors consider as a stopping criterion the feasibility of the last parametric non-linear program in particular by considering the complementarity constraint by the minimum component-wise. Table 3 provides our result with this criterion. We provide elementary statistics by considering the percentage of success for each set of parameter. A problem is considered solved in this case if criteria a) and b) are satisfied.

First, we see that the method NL is giving decent results. It is not a surprise as was pointed out in [25]. Practical implementation of relaxation methods would select the best choice of parameters so that we focus most of our attention to the line 'best'. In all cases, the relaxations manage to improve or at least equal the number of problem solved by NL. By using SNOPT, KS and butterfly with $t = r^{3/2}$ methods get 1% of improvement and with IPOPT the method butterfly with $t = r^{3/2}$ is the only one that attains 100%. The relaxation methods seem to give a significant improvement over NL with MINOS. In this case, it is clear that the butterfly methods benefit from the introduction of the parameter s and the method with $s = t, 2t = r$ is very competitive.

Solver snopt	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	97.03	97.03	98.02	97.03	97.03	98.02
average	97.03	95.02	94.71	95.39	93.89	94.88
worst	97.03	91.09	91.09	92.08	91.09	91.09
std	0	1.64	2.09	1.50	1.97	2.42
Solver minos	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	89.11	94.06	93.07	90.10	95.05	89.11
average	89.11	91.20	90.89	83.54	91.06	81.92
worst	89.11	87.13	87.13	77.23	86.14	76.24
std	0	1.50	1.44	2.81	2.15	2.89
Solver ipopt	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	98.02	99.01	98.02	99.01	98.02	100
average	98.02	98.16	96.38	94.03	93.89	94.79
worst	98.02	95.05	93.07	89.11	88.12	88.12
std	0	0.97	1.99	2.62	2.80	3.60

Table 3: Sensitivity analysis for MacMPEC test problems considering the feasibility of (MPCC). Results are percentage of success. best : percentage of success with the best set of parameter, worst : percentage of success with the worst set of parameter, average : average percentage of success among the distribution of (T, s), std : standard deviation

Our goal by solving (MPCC) is to compute a local minimum. The results using the local minimum criterion defined above as a measure of success are given in Table 4. Once again we provide percentage of success.

Solver snopt	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	92.08	94.06	96.04	96.04	97.03	96.04
average	92.08	90.78	91.17	92.08	90.04	92.33
worst	92.08	83.17	86.14	87.13	82.18	87.13
std	0	3.15	2.59	2.45	2.86	2.77
Solver minos	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	85.15	94.06	93.07	88.11	94.06	87.13
average	85.15	90.94	90.18	81.92	90.04	80.11
worst	85.15	87.13	86.14	76.23	85.15	74.26
std	0	1.50	1.62	2.65	2.31	2.95
Solver ipopt	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	91.09	93.07	93.07	94.06	93.07	94.06
average	91.09	91.82	89.84	89.05	88.80	89.02
worst	91.09	90.10	86.14	84.16	84.16	81.19
std	0	1.14	2.19	3.09	2.72	3.86

Table 4: Sensitivity analysis for MacMPEC test problems considering the optimality of (MPCC). Results are percentage of success. best : percentage of success with the best set of parameter, worst : percentage of success with the worst set of parameter, average : average percentage of success among the distribution of (T, s), std : standard deviation

In comparison with Table 3, this new criterion appears to be more selective. Independently of the solver, the relaxation methods with some correct choices of parameter provide improved results. Using SNOPT

as a solver, the methods KS and butterfly gives the highest number of results. The method butterfly with $t = r^{3/2}$ even improved the number of problem solved by SNOPT alone in average. In a similar way as in the previous experiment the butterfly method benefit of the introduction of the parameter s when using MINOS as a solver.

The relaxations (KDB) and (SU) have been discarded after preliminaries results. Indeed numerical difficulties can arise that are explained in the following two examples. In particular the following example illustrates one these difficulties for the relaxation (KDB). It is to be noted that both methods have received a special attention in [47] and [67] to solve the sub-problems that handle those issues.

Example 5.1 (KDB infeasible)**.**

$$\begin{aligned} \min_{x\in\mathbb{R}^2}\ & x_1 - x_2 \\ & x_1 \leq 0,\ x_2 \leq 0 \\ & 0 \leq x_1 \perp x_2 \geq 0 \end{aligned}$$

The feasible set of relaxation KDB is always empty for $t > 0$.

5.3 An Example of Numerical Difficulties

In this section, we illustrate the possible numerical difficulties that can arise by solving a (MPCC) with relaxation methods.

Example 5.2. *Consider the problem*

$$\begin{aligned} \min_{x\in\mathbb{R}^4}\ & \exp(-x_1^2 - x_2^2) + \exp(-x_3) \\ s.t.\ & x_3^2 \leq (x_1^2 + x_2^2 - 1)(x_1^2 + x_2^2 - 10) + x_4, \\ & x_1^2 + x_2^2 - 10 \leq 0, x_4^2 \leq 0, \\ & 0 \leq x_1^2 + x_2^2 - 1 \perp x_3(-x_1^2 - x_2^2 + 10) \geq 0. \end{aligned}$$

The feasible set is the union of two circles, $\{x \in \mathbb{R}^4 \mid x_3 = x_4 = 0,\ x_1^2 + x_2^2 = 1\}$ and $\{x \in \mathbb{R}^4 \mid x_3 = x_4 = 0,\ x_1^2 + x_2^2 = 10\}$. In this example, all the feasible points are local minima.

Let us now compute the stationary points of the problem. The gradient of MPCC-Lagrangian function equal to zero yields

$$\begin{aligned} -2\exp(-x_1^2 - x_2^2)x_1 - 2\lambda_1^g x_1((x_1^2 + x_2^2 - 10) + (x_1^2 + x_2^2 - 1)) + 2\lambda_2^g x_1 - 2\lambda^G x_1 + 2\lambda^H x_1 x_3 = 0, \\ -2\exp(-x_1^2 - x_2^2)x_2 - 2\lambda_1^g x_2((x_1^2 + x_2^2 - 10) + (x_1^2 + x_2^2 - 1)) + 2\lambda_2^g x_2 - 2\lambda^G x_2 + 2\lambda^H x_2 x_3 = 0, \\ -\exp(-x_3) + 2\lambda_1^g x_3 - \lambda^H(-x_1^2 - x_2^2 + 10) = 0, \\ -\lambda_1^g + 2\lambda_3^g x_4 = 0. \end{aligned}$$

It is clear that necessarily $x_3 = x_4 = 0$, thus $\lambda_1^g = 0$ and

$$\begin{aligned} -\exp(-x_1^2 - x_2^2)x_1 + \lambda_2^g x_1 - \lambda^G x_1 = 0, \\ -\exp(-x_1^2 - x_2^2)x_2 + \lambda_2^g x_2 - \lambda^G x_2 = 0, \\ -1 = \lambda^H(-x_1^2 - x_2^2 + 10). \end{aligned}$$

The third equality gives that $x_1^2 + x_2^2 \neq 10$, thus $\lambda_2^g = 0$. Furthermore, by the inequality constraints it is necessary that $x_1^2 + x_2^2 = 1$ and so either x_1 or x_2 is non-zero. It follows that $\lambda^H < 0$ and

$$-\exp(-1) = \lambda^G < 0.$$

$x_1^0 \backslash x_2^0$	0	1	2	3
0	o	C	C	M
1	C	C	C	M
2	C	C	M	M
3	M	M	M	M

Table 5: Sensitivity analysis depending on the initial point $(x_1^0, x_2^0, 0, 0)$ on Example 5.2 by using the butterfly relaxation method $t = r^{3/2}$ with $T = 0.5, s = 0.01$ and SNOPT as a non-linear solver. **Legend:** o: error, C: circle $x^2 + y^2 = 1$, M: circle $x^2 + y^2 = 10$.

relaxation	solver	output (last iter.)	last parameter	x_1	x_2	x_3	x_4
NL	SNOPT	401	.	0.1929	0.9812	0.0117	0.0001
SS		401	$2.5e-13$	0.1929	0.9812	0.0112	0.0001
KS		401	$5.0e-15$	0.1930	0.9811	0.0112	0.0001
$B_{(t=r^{3/2})}$		401	$5.0e-15$	0.1927	0.9812	0.0116	0.0001
NL	MINOS	0	.	0.7266	0.6870	0.0005	2.8389e-7
SS		0	$2.5e-9$	0.7266	0.6870	0.0007	5.3595e-7
KS		0	$5.0e-5$	0.7265	0.6870	0.0005	3.1903e-7
$B_{(t=r^{3/2})}$		0	$5.0e-5$	0.7266	0.6869	0.0005	3.5130e-7
NL	IPOPT	0	.	0.1819	0.9833	0.0100	9.9999e-5
SS		0	0.25	0.1961	0.9805	0.0100	9.9999e-5
KS		0	0.5	0.1961	0.9805	0.0100	9.9999e-5
$B_{(t=r^{3/2})}$		0	0.5	0.1961	0.9805	0.0100	9.9999e-5

Table 6: Example 5.2 with initial point $(0.1, 0.5)$. output 0 is a success and output 401 is iteration limit message.

To sum up, any point that satisfies $x_1^2 + x_2^2 = 1$ is C-stationary and is a local minimum, while any point that satisfies $x_1^2 + x_2^2 = 10$ is not stationary, despite the fact that it is a global minimum.

Up to this point, we may notice that the points that belong to the circle of centre 0 and radius $\sqrt{10}$ that are the global minima of the problem are sequentially M-stationary. Indeed, let $(x_1^k, x_2^k, x_3^k, x_4^k) = (0, \sqrt{10} - \frac{1}{k}, 0, 1/k)$, $\lambda^{H,k} = -\frac{1}{10-x_2^{k,2}} < 0$, $\lambda^{G,k} = 0$, $\lambda_1^{g,k} = \frac{-\exp(-x_1^{k,2}-x_2^{k,2})}{2\lambda_1^{g,k}(-x_2^{k,2}+\frac{11}{2})}$, $\lambda_2^{g,k} = 0$ and $2\lambda_3^{g,k} = k\lambda_1^{g,k}$.

We run Algorithm 2 with $T = 0,5$ and $s = 0,01$. Table 5 shows that the butterfly relaxation with $t = r^{3/2}$ may converge to both circles depending on the initial point. Note that for $(x_1^0, x_2^0) = (0, 0)$ the algorithm declares the problem infeasible. We do not give the results for other methods and other solvers here, but it has a similar behavior.

Those results may be surprising since it is proved that this method should converge to an M-stationary point and not less. So, in theory the algorithm should have some difficulties to compute Lagrange multiplier at this point. We run Algorithm 1 with methods NL, SS, KS and butterfly $t = r^{3/2}$ on this example. Results are presented in Table 6.

We see that independently of the solver all of the methods converge to a C-stationary point. In the cases of IPOPT and MINOS, the solvers exit with a success output and even more, they satisfy our local minimum criterion.

Those disturbing results are explained by Theorem 4.5 and related results in the literature that illustrate the fact that computing ϵ-stationary point may perturb the convergence properties of these methods. We also point out here that local minima of the problem are not M-stationary and so by Theorem 2.1 MPCC-GCQ does not hold at these points. Moreover, this example does not contradict the Theorem 4.5 since in

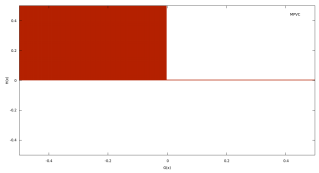

Figure 7: MPVC

particular MPCC-CRSC is not verified at any feasible point of the problem.

5.4 The Butterfly Relaxation for Mathematical Program with Vanishing Constraints

We consider the mathematical program with vanishing constraint

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{s.t. } g(x) \leq 0,\ h(x) = 0, \\ &\quad H(x) \geq 0, \\ &\quad G_i(x)H_i(x) \leq 0,\ i = 1,\dots,q, \end{aligned} \tag{MPVC}$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $G, H : \mathbb{R}^n \to \mathbb{R}^q$. This problem was first proposed by Achtziger in [4] motivated by applications in topology design and in mechanical structures problems. This problem can be reformulated as an (MPCC), however this runs at some constraint qualifications issues explaining why we need to propose specific numerical methods even so they are very close to methods for (MPCC). The feasible set of the vanishing constraint is given on Figure 7.

Relaxation methods from previous sections can be adapted to this case by considering the parametric non-linear program

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

Recent literature extend relaxation methods for (MPCC) to the problem (MPVC) : [5, 41] deal with the smooth method by Scheel & Scholtes, [3] consider a nonsmooth reformulation, [37] uses the local regularization from Steffensen & Ulbrich and finally [38] adapt the new paradigm of relaxation method with convergence to M-stationary point in particular the method by Kanzow & Schwartz.

Following the same path, we can extend the Butterfly relaxation method for MPVC :

$$\Phi^B_{(t,r)_i}(x) := G_i(x)(H_i(x) - t\theta_r(G_i(x))),\ i = 1,\dots,q$$

where θ is defined as before.

In order to validate this approach we run the method on an application of MPVC to truss topology optimization that was described in depth in chapter 9 of the monograph [36].

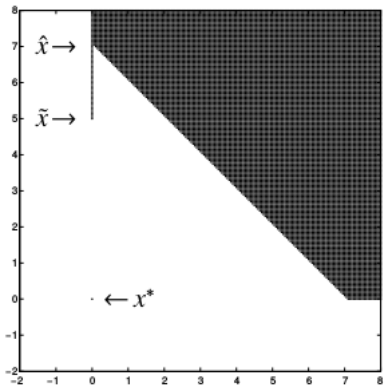

Figure 8: The feasible set of Example 5.3 from [36].

sovler	x^*	$f(x^*)$	last value of t
SNOPT	$(0,5)^T$	10	$0,5$
IPOPT	$(0,0)^T$	0	$4,67.10^{-4}$
MINOS	$(0,0)^T$	0	$0,5$

Table 7: Butterfly (MPVC) relaxation with $t = r^{3/2}$ on Example 5.3 with initial point $(6,6)^T$.

Example 5.3.

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \; & 4x_1 + 2x_2 \\ & x_1 \geq 0, \;\; x_2 \geq 0, \\ & (5\sqrt{2} - x_1 - x_2)x_1 \leq 0, \\ & (5 - x_1 - x_2)x_2 \leq 0. \end{aligned}$$

The feasible set of this example is given in Figure 8. As the geometry indicates, numerical methods based on feasible descent concepts generally converge to the point $\hat{x} = (0, 5\sqrt{2})^T$. The unique global solution to the problem is the point $x^* = (0,0)^T$. In practical application this point must be excluded by an additional constraint, and then the unique optimal solution to the problem is the point $\bar{x} = (0,5)^T$.

We run butterfly relaxation tailored to (MPVC) on Example 5.3 using an initial point inside the feasible domain $x^0 = (6,6)^T$. Results are presented in Table 7 with solvers SNOPT, IPOPT and MINOS. In two cases the butterfly method manages to converge to the global optimum and in the third case it converges to the point $(0,5)$ which is a local minimum.

6 Concluding Remarks

This paper proposes a new family of relaxation schemes for the mathematical program with complementarity constraints. We prove that these methods have the same theoretical properties as the best-known methods of [48] and [46] in the literature while extending them.

Convergence of the method is proved under the new and weak MPCC-CRSC condition. This new definition is also completed by the characterization of approximate stationary sequences that are proved to be bounded under this condition.

We provide a complete numerical study with remarks regarding the implementation as well as a numerical comparison with existing methods in the literature. These numerical experiments show that the new butterfly

schemes are very competitive. We also provide an example that illustrates some of the pitfalls that solvers may encounter while solving those degenerate non-linear programs.

The mathematical program with vanishing constraint is also a difficult non-linear program that need special care to handle. It appears as we shown that the butterfly schemes can also be adapted to this class of problems.

Further research, may focus on the two main difficulties regarding relaxation schemes that are the convergence of approximate stationary sequences and the existence of Lagrange multipliers. A discussion regarding the former problem a been initiated in [50] and appeal further discussion.

Acknowledgements

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

References

[1] Jean Abadie. *On the Kuhn-Tucker theorem.* Nonlinear Programming, 1967.

[2] Lina Abdallah and Mounir Haddou. An exact penalty approach for mathematical programs with equilibrium constraints. *Journal of Advances in Mathematics*, 9(8), 2013.

[3] Wolfgang Achtziger, Tim Hoheisel, and Christian Kanzow. A smoothing-regularization approach to mathematical programs with vanishing constraints. *Computational Optimization and Applications*, 55(3):733–767, 2013.

[4] Wolfgang Achtziger and Christian Kanzow. Mathematical programs with vanishing constraints: Optimality conditions and constraint qualifications. *Mathematical Programming*, 114(1):69–99, 2008.

[5] Wolfgang Achtziger, Christian Kanzow, and Tim Hoheisel. *On a relaxation method for mathematical programs with vanishing constraints*, volume 35. Inst. of Math., nov 2012.

[6] Sebastian Albrecht and Michael 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.

[7] Roberto Andreani, Gabriel Haeser, and José Màrio Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[8] Roberto Andreani, Gabriel Haeser, María Laura Schuverdt, and Paulo J.S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012.

[9] Roberto Andreani, José Mário Martinez, Alberto Ramos, and Paulo J.S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[10] Mihai Anitescu. On Using the Elastic Mode in Nonlinear Programming Approaches to Mathematical Programs with Complementarity Constraints. 15(4):1203–1236, 2005.

[11] Mokhtar S Bazaraa and Chitharanjan Marakada Shetty. *Foundations of optimization*, volume 122. Springer Science & Business Media, 2012.

[12] Maïtine Bergounioux and Mounir Haddou. A new relaxation method for a discrete image restoration problem. *Journal of Convex Analysis*, 17(3):421–435, 2010.

[13] J Frédéric Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

[14] Frank H Clarke. *Optimization and nonsmooth analysis*, volume 5. Siam, 1990.

[15] Zoumana Coulibaly and Dominique Orban. An ℓ_1 elastic interior-point method for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 22(1):187–211, jan 2012.

[16] Victor DeMiguel, Michael P. Friedlander, Francisco J. Nogales, and Stefan Scholtes. A two-sided relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 16(2):587–609, 2005.

[17] Stephan Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

[18] Stephan Dempe, Vyacheslav Kalashnikov, Gerardo A Pérez-Valdés, and Nataliya Kalashnykova. *Bilevel programming problems: theory, algorithms and applications to energy networks*. Springer, 2015.

[19] Michael L. Flegel. *Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints*. PhD thesis, University of Würzburg, 2005.

[20] Michael L. Flegel and Christian Kanzow. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005.

[21] Michael L. Flegel and Christian Kanzow. On M-stationary points for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 310(1):286–302, 2005.

[22] Michael L. Flegel and Christian Kanzow. On the guignard constraint qualification for mathematical programs with equilibrium constraints. *Optimization*, 54(6):517–534, 2005.

[23] Michael L. Flegel and Christian Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints*. Springer, 2006.

[24] Roger Fletcher and Sven Leyffer. Numerical experience with solving MPECs as NLPs. 2002.

[25] Roger Fletcher and Sven Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004.

[26] Roger Fletcher, Sven Leyffer, Danny Ralph, and Stefan Scholtes. Local convergence of sqp methods for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 17(1):259–286, 2006.

[27] Robert Fourer, David Gay, and Brian Kernighan. *Ampl*, volume 119. Boyd & Fraser, 1993.

[28] Masao Fukushima and Jong-Shi Pang. Convergence of a smoothing continuation method for mathematical progams with complementarity constraints. In *Ill-posed Variational Problems and Regularization Techniques*, pages 99–110. Springer, 1999.

[29] Philip E. Gill, Walter Murray, and Michael A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005.

[30] F. J. Gould and Jon W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, mar 1971.

[31] Monique Guignard. Generalized kuhn-tucker conditions for mathematical programming problems in a banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[32] Lei Guo, Gui-Hua Lin, and J. Ye Jane. Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 158(1):33–64, 2013.

[33] Lei Guo, Gui Hua Lin, and J. Ye Jane. Solving mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 166(1):234–256, 2015.

[34] Mounir Haddou. A new class of smoothing methods for mathematical programs with equilibrium constraints. *Pacific Journal of Optimization*, 5(1), 2009.

[35] Michael 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.

[36] Tim Hoheisel. *Mathematical programs with vanishing constraints*. PhD thesis, 2009.

[37] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints. *Optimization Methods and Software*, 27(3):483–512, 2012.

[38] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with vanishing constraints: a new regularization approach with strong convergence properties. *Optimization*, 61(6):619–636, 2012.

[39] Tim Hoheisel, Christian Kanzow, and Alexandra Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1-2):257–288, feb 2013.

[40] X. M. Hu and Danny Ralph. Convergence of a penalty method for mathematical programming with complementarity constraints. *Journal of Optimization Theory and Applications*, 123(2):365–390, nov 2004.

[41] Alexey F. Izmailov and Mikhail V. Solodov. Mathematical programs with vanishing constraints: Optimality conditions, sensitivity, and a relaxation method. *Journal of Optimization Theory and Applications*, 142(3):501–532, 2009.

[42] J. Ye Jane. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005.

[43] J. Ye Jane and X.Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, 22(4):977–997, 1997.

[44] Robert Janin. *Directional derivative of the marginal function in nonlinear programming*. Springer, 1984.

[45] Abdeslam Kadrani. *Nouvelles Régularisations pour l'optimisation avec contraintes de complémentarité*. PhD thesis, Université de Sherbrooke, 2008.

[46] Abdeslam Kadrani, Jean-Pierre Dussault, and Abdelhamid Benchakroun. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 20(1):78–103, 2009.

[47] Abdeslam Kadrani, Jean-Pierre Dussault, and Abdelhamid Benchakroun. A globally convergent algorithm for MPCC. *EURO Journal on Computational Optimization*, 3(3):263–296, 2015.

[48] Christian Kanzow and Alexandra 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.

[49] Christian Kanzow and Alexandra Schwartz. Convergence properties of the inexact Lin-Fukushima relaxation method for mathematical programs with complementarity constraints. *Computational Optimization and Applications*, 59(1-2):249–262, oct 2014.

[50] Christian Kanzow and Alexandra 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.

[51] Sven Leyffer. MacMPEC: AMPL collection of MPECs. *Argonne National Laboratory. Available at www. mcs. anl. gov/leyfier/MacMPEC*, 2000.

[52] Sven Leyffer, Gabriel López-Calva, and Jorge Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 17(1):52–77, 2006.

[53] Sven Leyffer and Todd S. Munson. A globally convergent filter method for MPECs. (April), 2009.

[54] Gui-Hua Lin and Masao Fukushima. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, 133(1-4):63–84, 2005.

[55] Xinwei Liu and Jie Sun. *Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints*, volume 101. 2004.

[56] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.

[57] Mathieu Marechal and Raphael Correa. A DC (Difference of Convex Functions) Approach Of The MPECS. *Optimization-Online.org*, X(0):1–18.

[58] M. Teresa T. Monteiro and José Filipe P. Meira. A penalty method and a regularization strategy to solve MPCC. *International Journal of Computer Mathematics*, 88(1):145–149, jan 2011.

[59] Bruce A Murtagh and Michael A Saunders. Minos 5.0 user's guide. Technical report, DTIC Document, 1983.

[60] Jirí 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.

[61] Jong-Shi Pang and Masao Fukushima. Complementarity constraint qualifications and simplified b-stationarity conditions for mathematical programs with equilibrium constraints3. 1998.

[62] Arvind U. Raghunathan and Lorenz T. Biegler. An interior point method for mathematical programs with complementarity constraints (MPCCs). *SIAM Journal on Optimization*, 15(3):720–750, 2005.

[63] Alberto Ramos. Mathematical Programms with Equilibrium Constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. Technical report, 2016.

[64] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[65] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[66] Stefan Scholtes and Michael Stöhr. How stringent is the linear independence assumption for mathematical programs with complementarity constraints? *Mathematics of Operations Research*, 26(4):851–863, 2001.

[67] Alexandra 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.

[68] Sonja Steffensen and Michael Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.

[69] Sonja Veelken. *A new relaxation scheme for mathematical programs with equilibrium constraints: Theory and numerical experience*. PhD thesis, Julius-Maximilians-Universität Würzburg, 2009.

[70] Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.