

The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints

Dussault, Jean-Pierre* Haddou, Mounir† Migot, Tang‡

Last update: august 2018

Abstract

We propose a new family of relaxation schemes for mathematical programs with complementarity constraints that extend the relaxation of Kadrani, Dussault, Benchakroun from 2009 and the one of Kanzow and Schwartz from 2011. We discuss the properties of the sequence of relaxed non-linear programs as well as stationary properties of limiting points. A sub-family of our relaxation schemes has the desired property of converging to a M-stationary point. A stronger convergence result is also proved in the affine case. We introduce new constraint qualifications, MPCC-CRSC and MPCC-GCRSC, to prove convergence of our method. In particular, the latter 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.

Keywords: non-linear programming ; MPCC ; MPEC ; relaxation methods ; stationary point ; constraint qualification ; CRSC

1 Introduction

We consider the Mathematical Program with Complementarity Constraints

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

with $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ and $G, H : \mathbb{R}^n \rightarrow \mathbb{R}^q$ 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 $i \in \{1, \dots, q\}$.

This problem has become an active subject in the literature in the last two decades. The wide variety of applications that can be cast as a MPCC is one of the reasons for this popularity. (MPCC) is clearly a non-linear programming problem and in general most of the functions involved in the formulation are non-convex.

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). This is partly due to the geometry of the complementarity constraint that always has an empty relative interior.

*Département d’Informatique, faculté des Sciences, Université de Sherbrooke, Canada. This research was partially supported by NSERC grant.

†IRMAR-Insa, Rennes, France

‡IRMAR-Insa, Rennes, France, e-mail: tangi.migot@gmail.com. This research was partially supported by a french grant from "l’Ecole des Docteurs de l’UBL" and "le Conseil Régional de Bretagne".

These issues have motivated the definition of enhanced constraint qualifications and optimality conditions for (MPCC) as in [7, 15, 16, 29] to cite some of the earliest research. In [8], Flegel and Kanzow provide an essential result that defines the right necessary optimality condition to (MPCC). This optimality condition is called M(Mordukhovich)-stationary condition. In view of the constraint qualifications issues that plague the (MPCC), the relaxation methods provide an intuitive answer. The complementarity constraint is relaxed using a parameter so that the new feasible domain is not thin anymore. It is assumed here that the classical constraints $g(x) \leq 0$ and $h(x) = 0$ are not more difficult to handle than the complementarity constraint. Finally, as the relaxing parameter is reduced, convergence to the feasible set of (MPCC) is obtained similarly to a homotopy technique.

These methods have been suggested in the literature back in 2000 by Scheel and Scholtes in [29, 30]. Their natural approach was later extended by Demiguel, Friedlander, Nogales and Scholtes in [4]. In [23], Lin and 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 [33], the authors consider a relaxation of the same type but only around the corner $G(x) = H(x) = 0$.

In the corresponding papers it has been shown that under classical conditions 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.

A new perspective for those schemes has been motivated in [17] providing an approximation scheme with convergence to M-stationary points by considering

$$(G_i(x) - t)(H_i(x) - t), \quad \forall i \in \{1, \dots, q\}. \quad (\text{KDB})$$

This is not a relaxation since the feasible domain of (MPCC) is not included in the feasible set of the sub-problems. The method has been extended has a relaxation method in [19] through an NCP function ϕ :

$$\phi(G_i(x) - t, H_i(x) - t), \quad \forall i \in \{1, \dots, q\}, \quad (\text{KS})$$

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:

$$\phi(G_i(x) - t_2 \theta_{t_1}(H_i(x)), H_i(x) - t_2 \theta_{t_1}(G_i(x))), \quad \forall i \in \{1, \dots, q\}. \quad (\text{Butterfly})$$

This new method handling two relaxing parameters allows to a non-linear perturbation of the domain.

The following example shows that the butterfly relaxation may improve relaxations from [17] and [19]. Indeed, it illustrates an example where their is no sequence of stationary point¹ that converges to a non-optimal point.

Example 1.1.

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

In this example, there are two stationary points: an S-stationary point $(1, 0)$ that is the global minimum and a M-stationary point $(0, 0)$, which is not a local minimum.² Unlike the relaxations (KDB) and (KS) where for $t_k = \frac{1}{k}$ a sequence $x^k = (t_k \ 2t_k)^T$, with $\lambda^{\Phi,k} = k$, may converge to $(0, 0)$ as k goes to infinity, there is no sequence of stationary point that converges to this undesirable point with the butterfly relaxation.

Our main contributions in this paper are the following:

1. We prove convergence of the butterfly relaxation scheme to A-stationary points in general and to M-stationary points for $t_{2,k} = o(t_{1,k})$.

¹Definition of stationary point of a non-linear program at the beginning of Section 2.1 page 3.

²Definitions of M- and S-stationarity are given in Definition 2.4 page 5.

2. We prove for the affine MPCC that the butterfly relaxation scheme converges to S-stationary points under MPCC-LICQ, thus generalizing the situation of Example 1.1.
3. We prove that the butterfly relaxation scheme computing approximate stationary points at each step converges to a M-stationary point assuming $t_{2,k} = o(t_{1,k})$ and $\epsilon_k = o(\max(G_i(x^k), H_i(x^k)))$.
4. We introduce the new MPCC-GCRSC³, which is the weakest known constraint qualification ensuring the existence of a bounded sequence of multipliers. Moreover, this new condition allows us to refine the proofs of convergence of the relaxation.
5. We provide extensive numerical results showing that the butterfly relaxation can efficiently solve the MPCC.

In Section 2, we introduce classical definitions and results from non-linear programming and MPCC theory. This section is completed by the definition of new constraint qualifications for MPCC called MPCC-CRSC and MPCC-GCRSC in Definition 2.8. In Section 3, we define the relaxation scheme with the new butterfly relaxation. 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 also generalizes the situation of Example 1.1 to illustrate a situation, where the non-linear perturbation allows to escape from undesirable points. Finally, in Section 5, we provide an extensive numerical study by giving details on the implementation, comparisons with other methods as well as an example that illustrates the numerical difficulties that might occur.

2 Preliminaries

(MPCC) is obviously a non-linear programming problem. Most of the 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 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

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

with $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Denote \mathcal{F} the feasible region of (NLP), the set of active indices $\mathcal{I}_g(x) := \{i \in \{1, \dots, p\} \mid g_i(x) = 0\}$. Let the Lagrangian $\mathcal{L}(x, \lambda)$ be $\mathcal{L}(x, \lambda) := f(x) + g(x)^T \lambda^g + h(x)^T \lambda^h$, where $\lambda = (\lambda^g, \lambda^h)$ is the vector of Lagrange multiplier.

We call a KKT point or a stationary point a couple (x, λ) with $x \in \mathcal{F}$ such that $\nabla_x \mathcal{L}(x, \lambda) = 0$, $\lambda^g \geq 0$ and $g(x)^T \lambda^g = 0$. We remind that the tangent cone of a set X at $x^* \in X$ is a closed cone defined by

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

Another useful tool for our study is the linearized cone of (NLP) at $x^* \in \mathcal{F}$ defined by

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

³A technical report discussing a condition similar to MPCC-GCRSC has been deposited on optimization-online.org [27], in a parallel but independent research.

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 main tool is the Karush-Kuhn-Tucker (KKT) conditions. Let x^* be a local minimum of (NLP) that satisfies a constraint qualification, then there exists a Lagrange multiplier λ^* such that (x^*, λ^*) is a KKT point of (NLP). Constraint qualifications are used to ensure the existence of the multiplier at x^* .

We now define some of the classical constraint qualifications. Note that there exists a wide variety of such notions and we define here those that are essential for our purpose.

Definition 2.1. Let $x^* \in \mathcal{F}$.

- (a) Linear Independence CQ (LICQ) holds at x^* if the family of gradients $\{\nabla g_i(x^*) \mid i \in \mathcal{I}_g(x^*)\}, \{\nabla h_i(x^*) \mid i = 1, \dots, m\}$ is linearly independent.
- (b) Constant Rank CQ (CRCQ) holds at x^* if there exists $\delta > 0$ such that for any subsets $I_1 \subseteq \mathcal{I}_g(x^*)$ and $I_2 \subseteq \{1, \dots, m\}$ the family of gradients $\{\nabla g_i(x) \mid i \in I_1\}, \{\nabla h_i(x) \mid i \in I_2\}$ has a constant rank for all $x \in \mathcal{B}_\delta(x^*)$.
- (c) Mangasarian-Fromovitz CQ (MFCQ) holds at x^* if the family of gradients $\{\nabla h_i(x^*) \mid i = 1, \dots, m\}$ is linearly independent and there exists a $d \in \mathbb{R}^n$ such that $\nabla g_i(x^*)^T d < 0$ ($i \in \mathcal{I}_g(x^*)$) and $\nabla h_i(x^*)^T d = 0$ ($i = 1, \dots, m$).
- (d) Constant Rank in the Subspace of Components (CRSC) holds at x^* if there exists $\delta > 0$ such that the family of gradients $\{\nabla g_i(x) \mid i \in J_-\}, \{\nabla h_i(x^*) \mid i = 1, \dots, m\}$ has the same rank for every $x \in \mathcal{B}_\delta(x^*)$, where $J_- := \{i \in \mathcal{I}_g(x^*) \mid -\nabla g_i(x^*) \in \mathcal{L}(x^*)^\circ\}$.

Remark 1. The definition of MFCQ given here is the most classical. It can be shown using some theorem of the alternative that this definition is equivalent to the family of active gradients being positively linearly independent.

In the last definition, C° denotes the polar of a cone C , defined as $C^\circ := \{z \in \mathbb{R}^n \mid z^T d \leq 0 \ \forall d \in C\}$. Constant rank of the subspace component, CRSC, was introduced recently in [1].

A local minimum is characterized by the fact that there is no feasible descent direction for the objective function of (NLP), that is $-\nabla f(x^*) \in \mathcal{T}_{\mathcal{F}}(x^*)^\circ$, where \mathcal{T}° denotes the polar cone of \mathcal{T} . On the other hand, the KKT conditions build ∇f using a linearization of the active constraints. In a classical way, we say that a point $x^* \in \mathcal{F}$ satisfies Guignard CQ if $\mathcal{T}_{\mathcal{F}}(x^*)^\circ = \mathcal{L}(x^*)^\circ$ and Abadie CQ if $\mathcal{T}_{\mathcal{F}}(x^*) = \mathcal{L}(x^*)$.

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. This has motivated the definition of the CCP condition in [2].

Definition 2.2. We say that a point $x^* \in \mathcal{F}$ satisfies the Cone-Continuity Property if the set-valued mapping $\mathbb{R}^n \ni x \Rightarrow K(x)$ such that

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

is outer semicontinuous at x^* , that is $\limsup_{x \rightarrow x^*} K(x) \subset K(x^*)$.

In this context, the outer limit is taken in the sense of Painlevé-Kuratowski corresponding to the Definition 5.4 given in [28]. It is to be noted here that $K(x)$ depends on x^* , since it considers only active constraints at x^* . Clearly, $K(x^*)$ is a closed convex cone and coincides with the polar linearized cone $\mathcal{L}(x^*)^\circ$. 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 [2] that CCP is strictly stronger than ACQ and weaker than CRSC.

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

Definition 2.3. Given a general non-linear program (NLP) and $\epsilon \geq 0$. We say that $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{p+m}$ is a ϵ -stationary point (or a ϵ -KKT point) if it satisfies

$$\begin{aligned}\|\nabla \mathcal{L}(x, \lambda)\|_\infty &\leq \epsilon, \|h(x)\|_\infty \leq \epsilon, \\ g_i(x) &\leq \epsilon, \lambda_i^g \geq 0, |\lambda_i^g g_i(x)| \leq \epsilon \forall i \in \{1, \dots, p\}.\end{aligned}$$

2.2 Mathematical Programs 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

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

In the sequel, we always consider these sets in x^* . In order to derive weaker optimality conditions, we consider an enhanced Lagrangian function. Let \mathcal{L}_{MPCC} be the generalized MPCC-Lagrangian function of (MPCC) such that

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

with $\lambda := (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$. It is clear that we cannot expect to compute usual KKT points since classical constraint qualifications, in general, do not hold, so we introduce weaker stationary concepts as in [29, 15].

Definition 2.4. A point $x^* \in \mathcal{Z}$ is said

- Weak (W)-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

$$\begin{aligned}\nabla_x \mathcal{L}_{MPCC}(x^*, \lambda) &= 0, \\ \lambda_i^g &= 0 (i \notin \mathcal{I}_g(x^*)), \lambda_{\mathcal{I}^{+0}}^G = 0, \lambda_{\mathcal{I}^{0+}}^H = 0.\end{aligned}$$

- Clarke (C)-stationary if x^* is W -stationary and $\forall i \in \mathcal{I}^{00}, \lambda_i^G \lambda_i^H \geq 0$.
- Alternatively or Abadie (A)-stationary if x^* is W -stationary and $\forall i \in \mathcal{I}^{00}, \lambda_i^G \geq 0$ or $\lambda_i^H \geq 0$.
- Mordukhovich (M)-stationary if x^* is W -stationary and $\forall i \in \mathcal{I}^{00}$, either $\lambda_i^G > 0, \lambda_i^H > 0$ or $\lambda_i^G \lambda_i^H = 0$.
- Strong (S)-stationary if x^* is W -stationary and $\forall i \in \mathcal{I}^{00}, \lambda_i^G \geq 0, \lambda_i^H \geq 0$.

Relations between these definitions are straightforward from the definitions.

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 \mathcal{L}_{MPCC} as in [29, 6]

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

This cone is no longer polyhedral and is not necessarily convex. However due to [6], one always has the following inclusions: $\mathcal{T}_{\mathcal{Z}}(x^*) \subseteq \mathcal{L}_{MPCC}(x^*) \subseteq \mathcal{L}(x^*)$.

Definition 2.5. Let $x^* \in \mathcal{Z}$. We say that MPCC-ACQ holds at x^* if $\mathcal{T}_{\mathcal{Z}}(x^*) = \mathcal{L}_{MPCC}(x^*)$ and MPCC-GCQ holds at x^* if $\mathcal{T}_{\mathcal{Z}}^\circ(x^*) = \mathcal{L}_{MPCC}(x^*)^\circ$.

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

Theorem 2.6 ([8]). A local minimum of (MPCC) that satisfies MPCC-GCQ or any stronger MPCC-CQ is a M-stationary point.

The polar of the MPCC-linearized cone is a key tool in the definition of constraint qualifications. It is, however, not trivial to compute. Therefore, we introduce the following:

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

Remark 2. When MPCC-GCQ holds at x^* , due to [24], one gets the following inclusion: $\mathcal{L}_{MPCC}(x^*)^\circ \subset \mathcal{P}_M(x^*)$.

We now introduce some constraint qualifications that will be used in the sequel. One of the main constraint qualifications used in the literature of (MPCC) is the MPCC-LICQ, see [31] for a discussion on this CQ. In a similar way we may extend CRCQ as in [12]. A condition that is similar was used in [19, 14] to prove convergence of relaxation methods for (MPCC). As pointed out in Theorem 2.6, the "correct" sign of the multiplier λ_i^G, λ_i^H for $i \in \mathcal{I}^{00}$ in the necessary optimality conditions for (MPCC) are the sign of M-stationary points. This motivates the definition of MPCC-GMFCQ that specializes the MPCC-MFCQ and the MPCC-LICQ by taking into account those signs of multipliers for $i \in \mathcal{I}^{00}$.

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

1. MPCC-LICQ holds at x^* if the only solution of

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

is the trivial solution.

2. MPCC-MFCQ holds at x^* if the only solution of (1) with $\lambda_i^g \geq 0$ ($i \in \mathcal{I}_g(x^*)$) is the trivial solution.

3. MPCC-GMFCQ holds at x^* if the only solution of (1) with $\lambda_i^g \geq 0$ ($i \in \mathcal{I}_g(x^*)$) and either $\lambda_i^G \lambda_i^H = 0$ or $\lambda_i^G > 0, \lambda_i^H > 0$ for all $i \in \mathcal{I}^{00}$ is the trivial solution.

Note here that MPCC-MFCQ and MPCC-GMFCQ have been defined using the alternative form of MFCQ mentioned in the Remark 1.

MPCC-LICQ and MPCC-MFCQ corresponds to the classical LICQ and MFCQ applied to the so-called Tighten NLP,

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } & g(x) \leq 0, h(x) = 0, G_i(x) = 0, H_i(x) \geq 0, (i \in \mathcal{I}^{0+}), \\ & G_i(x) \geq 0, H_i(x) = 0, (i \in \mathcal{I}^{+0}), \\ & G_i(x) = 0, H_i(x) = 0, (i \in \mathcal{I}^{00}). \end{aligned} \quad (\text{TNLP})$$

2.3 The New MPCC-GCRSC and MPCC-CRSC CQs

In a similar way as for MPCC-MFCQ and MPCC-GMFCQ, we extend the definition of CRSC constraint qualification to introduce the MPCC-CRSC and the MPCC-GCRSC, which are new in the MPCC literature. MPCC-CRSC is the specialization of CRSC applied to (TNLP).

Definition 2.8. Let $x^* \in \mathcal{Z}$.

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

$$\{\nabla g_i(x) \ (i \in \mathcal{I}_1), \ \nabla h_i(x) \ (i = 1, \dots, m), \ \nabla G_i(x) \ (i \in \mathcal{I}^{0+} \cup \mathcal{I}^{00}), \ \nabla H_i(x) \ (i \in \mathcal{I}^{00} \cup \mathcal{I}^{+0})\}$$

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

$$\mathcal{I}_1 := \{i \in \mathcal{I}_g(x^*) \mid -\nabla g_i(x^*) \in \mathcal{L}_{TNLP}(x^*)^\circ\}.$$

(b) MPCC-GCRSC holds at x^* if for any partition $\mathcal{I}_{++}^{00} \cup \mathcal{I}_{0-}^{00} \cup \mathcal{I}_{-0}^{00} = \mathcal{I}^{00}$ such that

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

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

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

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

$$\mathcal{I}_1 := \{i \in \mathcal{I}_g(x^*) \mid -\nabla g_i(x^*) \in \mathcal{P}_M(x^*)\},$$

$$\mathcal{I}_3 := \mathcal{I}^{0+} \cup \{i \in \mathcal{I}_{++}^{00} \mid \nabla G_i(x^*) \in \mathcal{P}_M(x^*)\} \cup \mathcal{I}_{-0}^{00},$$

$$\mathcal{I}_4 := \mathcal{I}^{+0} \cup \{i \in \mathcal{I}_{++}^{00} \mid \nabla H_i(x^*) \in \mathcal{P}_M(x^*)\} \cup \mathcal{I}_{0-}^{00}.$$

Furthermore, MPCC-GCRSC is weaker than MPCC-CRCQ. Indeed, MPCC-CRCQ requires that every family of linearly dependent gradients remains linearly dependent in some neighborhood, while the MPCC-GCRSC condition considers only the family of gradients that are linearly dependent with coefficients that have M-stationary signs.

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

Definition 2.9. We say that a feasible point x^* satisfies the MPCC-CCP if the set-valued mapping $\mathbb{R}^n \ni x \Rightarrow K_{MPCC}(x)$ such that

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

is outer semicontinuous at x^* .

This definition is motivated by sequential optimality conditions from [2] for non-linear programming and extended for (MPCC) in [26], where it has been proved to be a MPCC constraint qualification.

The following results give a characterization of some sequences that satisfy MPCC-CRSC and MPCC-GCRSC 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.

During the process of an iterative algorithm, we are interested in the study of accumulation points of sequences computed by the relaxation method. It is common to compute sequences that satisfy the following assumptions.

Assumption 1. Let $\{x^k\}$ and $0 \neq \{\lambda^k\} \in \mathbb{R}_+^p \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^q$ be two sequences that we assume satisfy: $x^k \rightarrow x^* \in \mathcal{Z}$,

$$\lim_{k \rightarrow \infty} \frac{\nabla \mathcal{L}_{MPCC}(x^k, \lambda^k)}{\|\lambda^k\|_\infty} = 0, \quad (\text{i})$$

and, denoting $\lambda^* := \lim_{k \rightarrow \infty} \lambda^k / \|\lambda^k\|_\infty$,

$$\lambda_i^{g,*} = 0 \ (i \notin \mathcal{I}_g(x^*)), \lambda_i^{G,*} = 0 \ (i \in \mathcal{I}^{+0}) \text{ and } \lambda_i^{H,*} = 0 \ (i \in \mathcal{I}^{0+}). \quad (\text{ii})$$

This condition may correspond to some kind of sequential optimality conditions.

The first step in our analysis is to prove that one sequence of multipliers satisfying Assumption 1 is bounded.

Theorem 2.10. Given two sequences $\{x^k\}, \{\lambda^k\}$ that satisfy Assumption 1 and suppose that MPCC-CRSC holds at x^* . Then, there exists a bounded sequence $\{\bar{\lambda}^k\}$ satisfying Assumption 1. Furthermore, x^* is a W -stationary point of the MPCC.

Proof. As in Assumption 1, we denote $\lambda^* := \lim_{k \rightarrow \infty} \lambda^k / \|\lambda^k\|_\infty$. First, we can build a sequence $\{\tilde{\lambda}^k\}$ such that $\tilde{\lambda}^k = \lambda^k$ for all indices except

$$\tilde{\lambda}_i^{g,k} = 0 \ (i \notin \text{supp}(\lambda^{g,*})), \tilde{\lambda}_i^{G,k} = 0 \ (i \notin \text{supp}(\lambda^{G,*})), \tilde{\lambda}_i^{H,k} = 0 \ (i \notin \text{supp}(\lambda^{H,*})).$$

Since $\{\lambda^k\}$ satisfies Assumption 1, it is clear that $\{\tilde{\lambda}^k\}$ also satisfies Assumption 1 as it can be noted that $\lim_{k \rightarrow \infty} \tilde{\lambda}^k / \|\tilde{\lambda}^k\|_\infty = \lambda^*$.

We first show that there exists a sequence, $\{\bar{\lambda}^k\}$, such that for all k the family of gradients of non-vanishing multipliers in the following equation are linearly independent:

$$w^k := \sum_{j \in \text{supp}(\lambda^{g,*})} \bar{\lambda}_j^{g,k} \nabla g_j(x^k) + \sum_{j=1}^m \bar{\lambda}_j^{h,k} \nabla h_j(x^k) - \sum_{j \in \text{supp}(\lambda^{G,*})} \bar{\lambda}_j^{G,k} \nabla G_j(x^k) - \sum_{j \in \text{supp}(\lambda^{H,*})} \bar{\lambda}_j^{H,k} \nabla H_j(x^k). \quad (2)$$

Note that by Assumption 1 the following inclusions hold: $\text{supp}(\lambda^{g,*}) \subset \mathcal{I}_g(x^*)$, $\text{supp}(\lambda^{G,*}) \subset \mathcal{I}^{+0} \cup \mathcal{I}^{00}$ and $\text{supp}(\lambda^{H,*}) \subset \mathcal{I}^{+0} \cup \mathcal{I}^{00}$.

Indeed, according to Lemma B.1, we can build a sequence of multipliers that satisfies (i) and (ii), such that the gradients corresponding to non-vanishing multipliers in equation (2) are linearly independent for all $k \in \mathbb{N}$. This may change the multipliers, but previously positive ones will stay at least non-negative and vanishing multipliers will remain zero. So, this new sequence satisfies Assumption 1.

We prove by contradiction that the sequence $\{\bar{\lambda}^k\}$ is bounded. If $\{\bar{\lambda}^k\}$ were not bounded, there would exist a subsequence such that

$$\frac{\bar{\lambda}^k}{\|\bar{\lambda}^k\|_\infty} \rightarrow \lambda^* \neq 0.$$

Here we consider a subsequence K , where the family of linearly independent gradients of non-vanishing multipliers is the same for all $k \in K$. Note that this can be done with no loss of generality, since there is a finite number of such subsequences and altogether they form a partition of the sequence.

Note that conditions (i) and (ii) give that $\lim_{k \rightarrow \infty} w^k / \|\bar{\lambda}^k\|_\infty = \lim_{k \rightarrow \infty} -\nabla f(x^k) / \|\bar{\lambda}^k\|_\infty = 0$. Thus, dividing by $\|\bar{\lambda}^k\|_\infty$ and passing to the limit in (2) yields

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

with $\lambda_j^{g,*} = 0$ for $j \notin \mathcal{I}_g(x^*)$, $\lambda_j^{G,*} = 0$ for $j \in \mathcal{I}^{+0}$ and $\lambda_j^{H,*} = 0$ for $j \in \mathcal{I}^{0+}$ by (ii).

It follows that the gradients with non-zero multipliers involved in the previous equation are linearly dependent. Moreover, for any j such that $\lambda_j^{g,*} > 0$ previous equation gives

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

It follows that $\langle -\nabla g_j(x^*), d \rangle \leq 0$ ⁴ for all $d \in \mathcal{L}_{TNLP}(x^*)$, and then $j \in \mathcal{I}_1$ in the definition of MPCC-CRSC.

MPCC-CRSC guarantees that these gradients remain linearly dependent in a whole neighborhood. This, however, is a contradiction to the linear independence of these gradients given by definition of the sequence $\{\bar{\lambda}^k\}$. Here, we used that for all k sufficiently large $\text{supp}(\lambda^k) \subseteq \text{supp}(\bar{\lambda}^k)$. Consequently, the sequence $\{\bar{\lambda}^k\}$ is bounded.

Finally, x^* is a W-stationary point with multipliers λ^* by Definition 2.4 and by passing to the limit in Assumption 1. \square

The following result is similar to Theorem 2.10 and focus on the case where the limit point is a M-stationary point.

Theorem 2.11. *Given two sequences $\{x^k\}, \{\lambda^k\}$ that satisfy Assumption 1 and suppose that MPCC-GCRSC holds at x^* . Furthermore, denoting $\lambda^* := \lim_{k \rightarrow \infty} \lambda^k / \|\lambda^k\|_\infty$, assume that for all $i \in \mathcal{I}^{00}$*

$$\lambda_i^{G,*} \lambda_i^{H,*} = 0 \text{ or } \lambda_i^{G,*} > 0, \lambda_i^{H,*} > 0. \quad (3)$$

Then, there exists a bounded sequence $\{\bar{\lambda}^k\}$ satisfying Assumption 1 and (3). Furthermore, x^ is a M-stationary point of the MPCC.*

Proof. The proof is completely similar to Theorem 2.10. Let $\{\bar{\lambda}^k\}$ be a sequence such that for all k the family of gradients of non-vanishing multipliers in (2) are linearly independent. Assuming that $\{\bar{\lambda}^k\}$ is not bounded, we can extract a subsequence such that $\frac{\bar{\lambda}^k}{\|\bar{\lambda}^k\|_\infty} \rightarrow \lambda^* \neq 0$.

Dividing by $\|\bar{\lambda}^k\|_\infty$ and passing to the limit in the equation (2) yields

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

with $\lambda_j^{g,*} = 0$ for $j \notin \mathcal{I}_g(x^*)$, $\lambda_j^{G,*} = 0$ for $j \in \mathcal{I}^{+0}$, $\lambda_j^{H,*} = 0$ for $j \in \mathcal{I}^{0+}$ and either $\lambda_j^{G,*} \lambda_j^{H,*} = 0$ or $\lambda_j^{G,*} > 0, \lambda_j^{H,*} > 0$ for $j \in \mathcal{I}^{00}$ by (ii) and (3).

The family of gradients considered in the definition of MPCC-GCRSC corresponds to the gradients with non-zero multipliers in the previous equation. Indeed, by linear dependence of the gradients at x^* any gradient whose multiplier is non-zero may be formulated as a linear combination of the other gradients.

Therefore, those gradients with non-vanishing multipliers belong to the polar of the M-linearized cone. MPCC-GCRSC guarantees that these gradients remain linearly dependent in a whole neighborhood, which contradicts the linear independence of these gradients given by definition of the sequence $\{\bar{\lambda}^k\}$. Thus, the sequence $\{\bar{\lambda}^k\}$ is bounded.

⁴Using the reformulation of $\nabla g_j(x^*)$ for $\lambda_j^{g,*} \neq 0$ and the specialization of \mathcal{L} to the (TNLP) yields to

$$\begin{aligned} & \langle -\nabla g_j(x^*), d \rangle \\ &= \sum_{i \in \mathcal{I}_g(x^*) \setminus j} \frac{\lambda_i^{g,*}}{\lambda_j^{g,*}} \langle \nabla g_i(x^*), d \rangle + \sum_{i=1}^m \frac{\lambda_i^{h,*}}{\lambda_j^{g,*}} \langle \nabla h_i(x^*), d \rangle - \sum_{i \in \mathcal{I}^{0+} \cup \mathcal{I}^{00}} \frac{\lambda_i^{G,*}}{\lambda_j^{g,*}} \langle \nabla G_i(x^*), d \rangle - \sum_{i \in \mathcal{I}^{+0} \cup \mathcal{I}^{00}} \frac{\lambda_i^{H,*}}{\lambda_j^{g,*}} \langle \nabla H_i(x^*), d \rangle \\ &= \sum_{i \in \mathcal{I}_g(x^*) \setminus j} \frac{\lambda_i^{g,*}}{\lambda_j^{g,*}} \langle \nabla g_i(x^*), d \rangle \leq 0 \end{aligned}$$

Finally, x^* is a M-stationary point with multipliers λ^* by Definition 2.4 and by passing to the limit in Assumption 1 and in (3). \square

We conclude this section by a consequence of Theorem 2.11 that states an essential result for this section, namely MPCC-GCRSC is a MPCC constraint qualification.

Corollary 2.12. *MPCC-GCRSC implies MPCC-CCP.*

Proof. We prove that a point x^* that satisfies MPCC-GCRSC satisfies the following relation $\limsup_{x \rightarrow x^*} K_{MPCC}(x) \subset K_{MPCC}(x^*)$.

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

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

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

The sequence $\{\lambda^k\}$ clearly satisfies the assumptions of Theorem 2.11. It follows that there exists an analogous bounded sequence and up to a subsequence we can extract a limit point λ^* . Consequently, 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 $j \in \mathcal{I}^{00}$.

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

We sum up this section in Figure 1 by giving the relationship between the various MPCC-CQ defined here. Note that MPCC-CRSC does not necessarily implies MPCC-GCRSC due to Remark 2 (page 6).

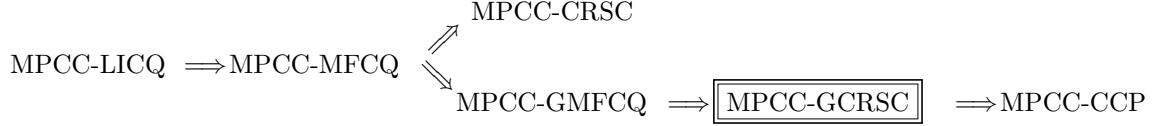


Figure 1: Relations between the MPCC constraint qualifications.

3 The Butterfly Relaxation Method

The focus of this paper is on relaxation methods to solve (MPCC). The sketch of such a method 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 $\{x^{k+1}\}$ should converge to a stationary point of (MPCC).

According to Theorem 2.6, our aim is to prove convergence to a M-stationary point of (MPCC). A motivation to consider relaxation methods is that the sequence of relaxed non-linear program may satisfy some constraint qualification and then are more tractable for classical non-linear methods.

We consider a family of continuously differentiable non-decreasing concave functions $\theta : \mathbb{R} \rightarrow]-\infty, 1]$ such that

$$\theta(0) = 0 \text{ and } \lim_{x \rightarrow \infty} \theta(x) = 1 \quad \forall x \in \mathbb{R}_{++}.$$

Then, we introduce $\theta_{t_1}(x) := \theta\left(\frac{x}{t_1}\right)$ for $t_1 > 0$ and complete in a smooth way for negative values by considering $\theta_{t_1}(z < 0) = z\theta'(0)/t_1$. Examples of such functions are $\theta_{t_1}^1(x) = \frac{x}{x+t_1}$ and $\theta_{t_1}^2(x) = 1 - \exp^{-\frac{x}{t_1}}$.

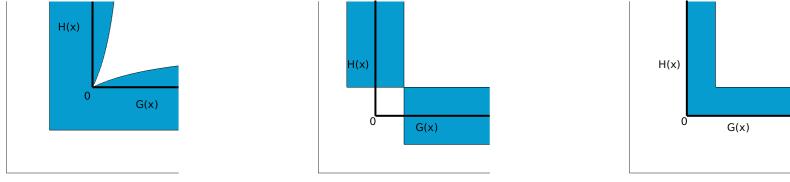


Figure 2: The feasible set of the butterfly relaxation, the approximation from [17] and the relaxation from [19].

Those functions have already been used in the context of complementarity constraints in [13]. Using this family of functions, we denote

$$F_{1i}(x; t_1, t_2) := H_i(x) - t_2 \theta_{t_1}(G_i(x)) \text{ and } F_{2i}(x; t_1, t_2) := G_i(x) - t_2 \theta_{t_1}(H_i(x)).$$

We propose a new family of relaxations with two positive parameters (t_1, t_2) defined such that for all $i \in \{1, \dots, q\}$

$$\Phi_i^B(G(x), H(x); t_1, t_2) := \begin{cases} F_{1i}(x; t_1, t_2) F_{2i}(x; t_1, t_2), & \text{if } F_{1i}(x; t_1, t_2) + F_{2i}(x; t_1, t_2) \geq 0, \\ -\frac{1}{2} (F_{1i}(x; t_1, t_2)^2 + F_{2i}(x; t_1, t_2)^2) & \text{otherwise.} \end{cases} \quad (\text{Bu.})$$

This definition already gives that $\Phi_i^B(G(x), H(x); t_1, t_2) = 0 \implies F_{1i}(x; t_1, t_2) + F_{2i}(x; t_1, t_2) \geq 0$.

This new relaxation uses two parameters t_1 and t_2 such that

$$t_2 \theta'(0) < t_1. \quad (4)$$

This condition ensures that the intersection point between the sets $\{x \in \mathbb{R}^n \mid F_1(x; t_1, t_2) = 0\}$ and $\{x \in \mathbb{R}^n \mid F_2(x; t_1, t_2) = 0\}$ is reduced to the origin. In other words, the two branches of the relaxation does not cross each other. A typical choice will be to take $t_2 = o(t_1)$ motivated by strong convergence properties as discussed in Section 4.1.

Most of the results presented in this article are only sensitive to the description of the constraint at its boundary and more general formulations of this regularization can be found for instance in [24].

We also add a regularization of the positivity constraints parametrized by \bar{t} .

This method is similar to the methods (KDB) from [17] and (KS) from [19] in the sense that they can also be written as a product of two functions. The main difference is that handling two parameters allows bringing the two "wings" of the relaxation closer. This observation motivated to consider algorithmic properties of a class of relaxation methods in a recent working paper [5]. A comparison of the feasible set of these relaxations can be seen in Figure 2.

We now introduce some notations that will be extensively used in the sequel. Since the butterfly relaxation uses two parameters we denote $t := (t_1, t_2)$ to simplify the notation and by extension $t_k := (t_{1,k}, t_{2,k})$.

Let $\mathcal{X}_{t, \bar{t}}^B$ be the feasible set of $R_{t, \bar{t}}^B$, which corresponds to the non-linear program related to the butterfly relaxation of the complementarity constraints defined in (Bu.), that is

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t } x \in \mathcal{X}_{t, \bar{t}}^B, \quad (R_{t, \bar{t}}^B)$$

with

$$\mathcal{X}_{t, \bar{t}}^B := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0, G(x) \geq -\bar{t}e, H(x) \geq -\bar{t}e, \Phi^B(G(x), H(x); t) \leq 0\},$$

where e denotes the vector of all ones.

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

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

Several relations between these sets follow directly from the definition of the relaxation. For instance, it holds that

$$\mathcal{I}_G \cap \mathcal{I}_{GH} = \mathcal{I}_H \cap \mathcal{I}_{GH} = \emptyset.$$

The following two lemmas give more insights on the relaxation.

Lemma 3.1. *Let $x \in \mathcal{X}_{t, \bar{t}}^B$, then it is true for the relaxation (Bu.) that:*

- (a) $\{i \in \mathcal{I}_{GH}(x; t) \mid F_{1i}(x; t) = 0, F_{2i}(x; t) < 0\} = \{i \in \mathcal{I}_{GH}(x; t) \mid F_{1i}(x; t) < 0, F_{2i}(x; t) = 0\} = \emptyset$;
- (b) $i \in \mathcal{I}_{GH}(x; t) \implies G_i(x) \geq 0, H_i(x) \geq 0$.

Proof. Case (a) is direct considering that $\Phi_i^B(G(x), H(x); t_1, t_2) \neq 0$ for $F_{1i}(x; t) + F_{2i}(x; t) < 0$.

By symmetry of the relaxation it is sufficient to assume that $F_{1i}(x; t) = H_i(x) - t_2 \theta_{t_1}(G_i(x)) = 0$ for some $i = 1, \dots, q$. Then, by definition of $F_{2i}(x; t)$ it holds that

$$F_{2i}(x; t) = G_i(x) - t_2 \theta_{t_1}(H_i(x)) = G_i(x) - t_2 \theta_{t_1}(t_2 \theta_{t_1}(G_i(x))),$$

so $G_i(x) \geq 0$ since in the other case by definition of the function θ it would follow that

$$F_{2i}(x; t) = G_i(x)(1 - (\theta'(0)t_2/t_1)^2),$$

which would be negative if $G_i(x) < 0$ using (4). Finally, $G_i(x) \geq 0$ implies that $H_i(x) \geq 0$ since $F_{1i}(x; t) = 0$. \square

The following lemma state two of the key features of the relaxation.

Lemma 3.2. *$\mathcal{X}_{t, \bar{t}}^B$ satisfy the following properties:*

1. $\mathcal{X}_{0,0}^B = \mathcal{Z}$;
2. $\cap_{t, \bar{t} \geq 0} \mathcal{X}_{t, \bar{t}}^B = \mathcal{Z}$.

If the feasible set of the (MPCC) is non-empty, then the feasible set of the relaxed sub-problems are also non-empty for all $t \geq 0$. If for some parameter $t \geq 0$ the set $\mathcal{X}_{t, \bar{t}}^B$ is empty, then it implies that \mathcal{Z} is empty. Finally, if a local minimum of $R_{t, \bar{t}}^B$ already belongs to \mathcal{Z} , then it is a local minimum of the (MPCC).

We focus in the sequel on the properties of these new relaxation schemes. We prove that the method converges to an A-stationary point in Theorem 4.1 and to a M-stationary point, Theorem 4.2, with some relation between the sequences $\{t_{2,k}\}$ and $\{t_{1,k}\}$. Furthermore, we prove in the affine case convergence to S-stationary point under MPCC-LICQ (Theorem 4.3).

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 the Guignard CQ as proved in Theorem 4.4.

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 statements regarding convergence and regularity properties of the butterfly relaxation, we provide some useful results on the asymptotic behavior of functions θ_{t_1} and $\Phi^B(G(x), H(x); t)$. Direct computation gives the gradient of $\Phi^B(G(x), H(x); t)$ in the following lemma.

Lemma 3.3. For all $i \in \{1, \dots, q\}$, the gradient of $\Phi_i^B(G(x), H(x); t)$ w.r.t. x for the relaxation (Bu.) is given by

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

The following lemma illustrates the behavior of functions θ_{t_1} and their derivatives. Its proof is a direct consequence of Lemma 1.4.5 in [24].

Lemma 3.4. Given two sequences $\{t_{1,k}\}$ and $\{t_{2,k}\}$, which converge to 0 as k goes to infinity and $\forall k \in \mathbb{N}, (t_{1,k}, t_{2,k}) \in \mathbb{R}_{++}^2$. It holds true that

$$\lim_{k \rightarrow \infty} t_{2,k} \theta'_{t_{1,k}}(z) = 0 \quad \forall z \in \mathbb{R}_{++}.$$

Furthermore, assuming that $t_{2,k} = o(t_{1,k})$, yields to

$$\lim_{k \rightarrow \infty} t_{2,k} \theta'_{t_{1,k}}(z) = 0 \quad \forall z \in \mathbb{R}_+.$$

4 Theoretical Properties

The study of theoretical properties of the butterfly relaxation method 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 approximate stationary points.

4.1 Convergence

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

We prove in Theorem 4.1 that the butterfly relaxation converges to an A-stationary point. This result is improved to a convergence to M-stationary points for some choices on the parameters t_2 and t_1 in Theorem 4.2.

Finally, we prove convergence to S-stationary point in a specific situation in Theorem 4.3. To the best of our knowledge this kind of result without second order condition is new from the literature and allows the butterfly relaxation to escape from spurious points.

Theorem 4.1. Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying (4) and decreasing to zero as k goes to infinity. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ 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_{t_k, \bar{t}_k}^B for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-CRSC holds at x^* . Then, x^* is an A-stationary point.

Proof. First, we identify the expressions of the multipliers of the complementarity constraint in Definition 2.4 through the stationary points of R_{t_k, \bar{t}_k}^B . The representation of $\nabla \Phi^B$ immediately gives $\nabla \Phi_i^B(G(x^k), H(x^k); t_k) = 0, \forall i \in \mathcal{I}_{GH}^{00}(x^k; t_k)$ for all $k \in \mathbb{N}$. Thus, we can write

$$-\nabla f(x^k) = \sum_{i=1}^p \lambda_i^{g,k} \nabla g_i(x^k) + \sum_{i=1}^m \lambda_i^{h,k} \nabla h_i(x^k) - \sum_{i=1}^q \lambda_i^{G,k} \nabla G_i(x^k) - \sum_{i=1}^q \lambda_i^{H,k} \nabla H_i(x^k), \quad (5)$$

where $\lambda^{g,k} = \eta^{g,k}$, $\lambda^{h,k} = \eta^{h,k}$ and

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

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

Notice that $i \in \{i = 1, \dots, q \mid F_{1i}(x^k; t_k) = 0\}$ implies that $i \in \mathcal{I}_{GH}^{0+}(x^k; t_k) \cup \mathcal{I}_{GH}^{00}(x^k; t_k)$ or symmetrically $i \in \{i \mid F_{2i}(x^k; t_k) = 0\}$ implies that $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k) \cup \mathcal{I}_{GH}^{00}(x^k; t_k)$ by concavity and $t_{2,k} \theta'(0) \leq t_{1,k}$ for all $k \in \mathbb{N}$.

First, we prove that there exists a bounded sequence $\{\bar{\lambda}^k\}$ using Theorem 2.10. Thus, we check that the sequence $\{\lambda^k\}$ verifies Assumption 1. Denote $\|\lambda^k\|_\infty := \|\lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}\|_\infty$. Using the definition of $\lambda^{G,k}$ and $\lambda^{H,k}$ in (5) and since by (4) it holds that $t_k \theta'_{t_{1,k}}(G_i(x^k)) \leq t_k \theta'_{t_{1,k}}(0) < 1$ for all $i \in \mathcal{I}_{GH}(x^k; t_k)$, it can be observed that⁵

$$\|\lambda^k\|_\infty = \|\eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k} \circ F_2(x^k; t_k), \eta^{\Phi,k} \circ F_1(x^k; t_k)\|_\infty, \quad (6)$$

where \circ denotes the componentwise product of two vectors.

- By (5), assumption (i) is verified.
- Let us now verify condition (ii).

By definition of $\{\lambda^{g,k}\}$ it holds that $\mathcal{I}_g(x^k) \subset \mathcal{I}_g(x^*)$ and so $\forall i \notin \mathcal{I}_g(x^*) \lim_{k \rightarrow \infty} \lambda_i^{g,k} = 0$.

Let $i \in \mathcal{I}^{+0}$, we verify that $\lim_{k \rightarrow \infty} \frac{\lambda_i^{G,k}}{\|\lambda^k\|_\infty} = 0$. The case $\forall i \in \mathcal{I}^{0+} \lim_{k \rightarrow \infty} \frac{\lambda_i^{H,k}}{\|\lambda^k\|_\infty} = 0$ will follow symmetrically.

Notice that $\eta_i^{G,k} = 0$, since stationary condition implies that $\eta_i^{G,k}(G_i(x^k) + \bar{t}_k) = 0$ and $G_i(x^k) \rightarrow G_i(x^*) > 0$.

Assume by contradiction that $\lim_{k \rightarrow \infty} \frac{\lambda_i^{G,k}}{\|\lambda^k\|_\infty} \neq 0$. By definition of $\lambda^{G,k}$ and since $\eta_i^{G,k} = 0$, this implies that $\eta_i^{\Phi,k} > 0$. As a consequence $i \in \mathcal{I}_{GH}(x^k; t_k)$ and in particular $i \in \mathcal{I}_{GH}^{0+}(x^k; t_k)$ or $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k)$. It cannot be $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k)$, since it would follow that $F_{2i}(x^k; t_k) = G_i(x^k) - t_{2,k} \theta'_{t_{1,k}}(H_i(x^k)) = 0$, which would contradict $G_i(x^k) \rightarrow G_i(x^*) > 0$. Therefore $i \in \mathcal{I}_{GH}^{0+}(x^k; t_k)$, which yields

$$\lambda_i^{G,k} = \eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) F_{2i}(x^k; t_k).$$

Moreover, thanks to (6) $\|\lambda^k\|_\infty \geq |\eta_i^{\Phi,k} F_{2i}(x^k; t_k)|$, thus

$$\frac{\lambda_i^{G,k}}{\|\lambda^k\|_\infty} \leq \frac{\eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) F_{2i}(x^k; t_k)}{|\eta_i^{\Phi,k} F_{2i}(x^k; t_k)|} = t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) \rightarrow 0$$

since $G_i(x^k) \rightarrow G_i(x^*) > 0$ by Lemma 3.4.

⁵For indices $i \in \mathcal{I}_{GH}^{0+}(x^k; t_k)$ (symmetry for indices $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k)$), then $\lambda_i^{G,k} = \eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) F_{2i}(x^k; t_k)$ and $\lambda_i^{H,k} = \eta_i^{\Phi,k} F_{2i}(x^k; t_k)$. Therefore, considering that $t_k \theta'_{t_{1,k}}(G_i(x^k)) < 1$, we get $\lambda_i^{G,k} < \lambda_i^{H,k}$. All in all the infinite norm is not obtained at these components.

Since MPCC-CRSC holds at x^* , by Theorem 2.10 there exists a bounded sequence $\{\bar{\lambda}^k\}$ satisfying Assumption 1. Furthermore, x^* is a W-stationary point of the MPCC.

Finally, let us now verify that x^* is an A-stationary point by computing the multipliers for indices $i \in \mathcal{I}^{00}$. Beforehand, by the proof of Theorem 2.10, the sequence $\{\bar{\lambda}^k\}$ shares the same sign pattern as the sequence $\{\lambda^k\}$ defined in (5). So, it suffices to show that $\lambda^* := \lim_{k \rightarrow \infty} \lambda^k / \|\lambda^k\|_\infty$ has the A-stationary signs for $i \in \mathcal{I}^{00}$.

Let $i \in \mathcal{I}^{00}$. Assume without loss of generality that $\lambda_i^{G,*} < 0$ (the other case will be similar by symmetry) and we show that $\lambda_i^{H,*} \geq 0$. $\lambda_i^{G,*} < 0$ implies that $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k)$ for k sufficiently large by definition of $\lambda_i^{G,k}$. So, $\lambda_i^{H,k} = \eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(H_i(x^k)) F_{1i}(x^k; t_k)$, which is non-negative.

So x^* is an A-stationary point. \square

The following example shows that the result of Theorem 4.1 is sharp since convergence cannot be ensured assuming only that MPCC-GCRSC or even MPCC-GMFCQ holds at the limit point.

Example 4.1. Consider the following two-dimensional example

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

MPCC-GMFCQ holds at $(0, 0)^T$. However, MPCC-CRSC obviously fails to hold at this point. The point $(0, 0)^T$ is even not a W-stationary point.

In this case, there exists a sequence of stationary points of the relaxation such that $\{x^k\}$ converges to the origin. Given a sequence $\{x^k\}$, with $\{1\} \in \mathcal{I}_{GH}(x^k; t_k)$, such that $x^k \rightarrow (0, 0)^T$ then $\eta^{G,k} = \eta^{H,k} = 0$ and we can choose $\eta^{\Phi,k}$ that satisfies

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

The sequence $\{x^k\}$ converges to an undesirable point.

The result of the Theorem 4.1 can be tightened if we consider a particular choice of parameter. It is an essential result, since it shows that a subfamily of the butterfly relaxation has the desired property to converge to a M-stationary point.

Theorem 4.2. Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying (4) and decreasing to zero as k goes to infinity. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ 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_{t_k, \bar{t}_k}^B for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-GCRSC holds at x^* . If $t_{2,k} = o(t_{1,k})$, then, x^* is an M-stationary point.

Proof. We use Theorem 2.11 to show the result. In the beginning of the proof of Theorem 4.1 (equation (5)) we define a sequence $\lambda^k := \{\lambda^{g,k}, \lambda^{h,k}, \lambda^{G,k}, \lambda^{H,k}\}$ satisfying Assumption 1. Let us now check that for $t_{2,k} = o(t_{1,k})$ this sequence also verifies (3).

Let $i \in \mathcal{I}^{00}$. The sequence $\{\lambda^k\} / \|\lambda^k\|_\infty$ is bounded and converges, up to a subsequence, to some limit denoted λ^* .

We check that $\lambda_i^{G,*} < 0 \implies \lambda_i^{H,*} = 0$ (the other case will be similar by symmetry). $\lambda_i^{G,*} < 0$ implies that $i \in \mathcal{I}_{GH}^{+0}(x^k; t_k)$ for k sufficiently large by definition of $\lambda_i^{G,k}$. So, $\lambda_i^{H,k} = \eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(H_i(x^k)) F_{1i}(x^k; t_k)$. Moreover, $\lim_{k \rightarrow \infty} t_{2,k} \theta'_{t_{1,k}}(H_i(x^k)) = 0$ by Lemma 3.4 with $t_{2,k} = o(t_{1,k})$ and $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{1i}(x^k; t_k) / \|\lambda^k\|_\infty = \lambda^{G,*} < \infty$. Thus, $\lim_{k \rightarrow \infty} \lambda_i^{H,k} / \|\lambda^k\|_\infty = 0$. So, the sequence $\{\lambda^k\}$ verifies assumption (3).

Finally, since MPCC-GCRSC holds at x^* , Theorem 2.11 gives the result. \square

The following example shows that this result is sharp, since it illustrates a situation where MPCC-GCRSC does not hold and the method converges to an undesirable W-stationary point. This phenomenon only happens if the sequence of multipliers defined in (5) is unbounded.

Example 4.2. Consider the problem

$$\min_{x \in \mathbb{R}^2} x_2^2 \text{ s.t. } 0 \leq x_1^2 + x_1 + x_2^2 \geq 0.$$

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

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

Consider a sequence such that for $(t_{1,k}, t_{2,k})$ sufficiently small $F_2(x^k; t_k) = 0$ and

$$x_1^k = t_{2,k} \theta'_{t_{1,k}}(x_1^k + a^2), \quad x_2^k = a, \quad \eta^{\Phi,k} F_{1i}(x^k; t_k) = \frac{1}{-t_{2,k} \theta'_{t_{1,k}}(x_1^k + a^2)}.$$

Obviously, the sequence x^k goes to $x^* = (0, a \neq 0)^T$, which is not a W-stationary point. Indeed, we have

$$\lambda^{G,k} = \frac{1}{t_{2,k} \theta'_{t_{1,k}}(x_1^k + a^2)} \rightarrow \infty \text{ and } \lambda^{H,k} = -1 \neq 0.$$

The following result motivated by Example 1.1 shows that the butterfly relaxation may improve its behaviour in some specific cases. Example 1.1 also shows that this cannot be expected with the other relaxations.

Theorem 4.3. Assume that f, g, h, G, H are affine functions. Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying (4) and decreasing to zero as k goes to infinity. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ 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_{t_k, \bar{t}_k}^B for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-LICQ holds at x^* . If $t_{2,k} = o(t_{1,k})$, then, x^* is an S-stationary point.

Proof. Theorem 4.2 already proves that x^* is an M-stationary point. Assume by contradiction that x^* is not an S-stationary point. Then, it holds that this point cannot be a stationary point of R_{t_k, \bar{t}_k}^B .

We already mention in the proof of Theorem 4.1 that for all k it holds

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

where we omit the dependence in k in the expression of the gradients, since they are constant by linear/affine assumption. Now, by continuity linear independence of these gradients holds in a neighbourhood of x^* . So, we get finite convergence of the λ^k , and for k sufficiently large it holds

$$\lambda^{g,k} = \lambda^{g,\infty}, \quad \lambda^{h,k} = \lambda^{h,\infty}, \quad \lambda^{G,k} = \lambda^{G,\infty}, \quad \lambda^{H,k} = \lambda^{H,\infty}. \quad (7)$$

Let $i \in \mathcal{I}^{00} \cap \text{supp}(\eta^{\Phi,\infty})$, where we remind that $\text{supp}(\eta^{\Phi,\infty}) \subset \text{supp}(\eta^{\Phi,k}) \subset \mathcal{I}_{GH}(x^k; t_k)$. If no such index exists, then for all k sufficiently large $\eta^{\Phi,k}$ is zero and x^* is S-stationary.

By stationarity assumption on x^* , we assume that $\lambda_i^{G,\infty} < 0$ (the case $\lambda_i^{H,\infty}$ will be symmetrical). It implies that $i \in \mathcal{I}_{GH}^{+0}$ by definition of the multipliers in (5) and so

$$\lambda_i^{G,k} = \eta_i^{\Phi,k} F_{1i}(x^k; t_k) \text{ and } \lambda_i^{H,k} = \eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(H_i(x^k)) F_{1i}(x^k; t_k).$$

We prove in previous convergence result that $\lambda^{G,k}$ and $\lambda^{H,k}$ have bounded limits, so by Lemma 3.4 with $t_{2,k} = o(t_{1,k})$ we have $\lim_{k \rightarrow \infty} \lambda_i^{H,k} = 0$. By (7), we get $\eta_i^{\Phi,k} = 0$ for all k sufficiently large, which contradicts $\lambda_i^{G,\infty} < 0$. \square

4.2 Existence of Lagrange Multipliers for the Relaxed Sub-Problems

In this part, we study some 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 relaxation satisfies Guignard CQ as stated in Theorem 4.4, which is equivalent in terms of regularity to the relaxation (KS).

Theorem 4.4. *Let $x^* \in \mathcal{Z}$, satisfying MPCC-LICQ. Then, there exists $t^* > 0$ and a neighborhood $U(x^*)$ of x^* such that:*

$$\forall t \in (0, t^*] : x \in U(x^*) \cap \mathcal{X}_{t, \bar{t}}^B \implies \text{standard GCQ holds at } x \text{ for } R_{t, \bar{t}}^B.$$

Proof. Let $x \in U(x^*) \cap \mathcal{X}_{t, \bar{t}}^B$. We know that $\mathcal{L}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ \subseteq \mathcal{T}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ$. So, it is sufficient to show the converse inclusion.

The linearized cone of $R_{t, \bar{t}}^B$ is given by

$$\begin{aligned} \mathcal{L}_{\mathcal{X}_{t, \bar{t}}^B}(x) = \{d \in \mathbb{R}^n \mid & \nabla g_i(x)^T d \leq 0 \ (i \in \mathcal{I}_g(x)), \ \nabla h_i(x)^T d = 0 \ (i = 1, \dots, m), \\ & \nabla G_i(x)^T d \geq 0 \ (i \in \mathcal{I}_G(x; \bar{t})), \ \nabla H_i(x)^T d \geq 0 \ (i \in \mathcal{I}_H(x; \bar{t})), \\ & \nabla \Phi_i^B(G(x), H(x); t)^T d \leq 0 \ (i \in \mathcal{I}_{GH}^{0+}(x; t) \cup \mathcal{I}_{GH}^{+0}(x; t))\}, \end{aligned}$$

using that $\nabla \Phi_i^B(G(x), H(x); t) = 0$ for all $i \in \mathcal{I}_{GH}^{00}(x, t)$.

Let us compute the polar of the tangent cone. Consider the following set of non-linear constraints parametrized by $z \in \mathcal{X}_{t, \bar{t}}^B$ and a partition (I, I^c, I^-) of $\mathcal{I}_{GH}^{00}(z; t)$ ⁶, defined by

$$\begin{aligned} S_{(I, I^c, I^-)}(z) := \{x \in \mathbb{R}^n \mid & g(x) \leq 0, h(x) = 0, G(x) \geq -\bar{t}e, H(x) \geq -\bar{t}e, \\ & \Phi_i^B(G(x), H(x); t) \leq 0, i \notin \mathcal{I}_{GH}^{00}(z; t), \\ & F_{1i}(x; t) \leq 0, F_{2i}(x; t) \geq 0, i \in I, \\ & F_{1i}(x; t) \geq 0, F_{2i}(x; t) \leq 0, i \in I^c, \\ & F_{1i}(x; t) \leq 0, F_{2i}(x; t) \leq 0, i \in I^-\}. \end{aligned} \tag{8}$$

Since $z \in \mathcal{X}_{t, \bar{t}}^B$, it is obvious that $z \in S_{(I, I^c, I^-)}(z)$.

By construction of $U(x^*)$ and t^* , the gradients $\{\nabla g_i(x^*) \ (i \in \mathcal{I}_g(x^*)), \nabla h_i(x^*) \ (i = 1, \dots, m), \nabla G_i(x^*) \ (i \in \mathcal{I}^{00} \cup \mathcal{I}^{0+}), \nabla H_i(x^*) \ (i \in \mathcal{I}^{+0} \cup \mathcal{I}^{00})\}$ remain linearly independent for all $x \in U(x^*)$ by continuity of the gradients and we have

$$\begin{aligned} \mathcal{I}_g(x) \subseteq \mathcal{I}_g(x^*), \ \mathcal{I}_G(x; \bar{t}) \subseteq \mathcal{I}^{00} \cup \mathcal{I}^{0+}, \ \mathcal{I}_H(x; \bar{t}) \subseteq \mathcal{I}^{+0} \cup \mathcal{I}^{00}, \\ \mathcal{I}_{GH}^{00}(x; t) \cup \mathcal{I}_{GH}^{+0}(x; t) \subseteq \mathcal{I}^{00} \cup \mathcal{I}^{0+}, \\ \mathcal{I}_{GH}^{00}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t) \subseteq \mathcal{I}^{+0} \cup \mathcal{I}^{00}. \end{aligned} \tag{9}$$

Therefore, by Lemma A.1, LICQ holds for (8) at x . Furthermore, by Lemma B.2 and since LICQ in particular implies Abadie CQ it follows that

$$\mathcal{T}_{\mathcal{X}_{t, \bar{t}}^B}(x) = \bigcup_{\forall (I, I^c, I^-)} \mathcal{T}_{S_{(I, I^c, I^-)}(x)}(x) = \bigcup_{\forall (I, I^c, I^-)} \mathcal{L}_{S_{(I, I^c, I^-)}(x)}(x).$$

By [3, Theorem 3.1.9], passing to the polar, we get

$$\mathcal{T}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ = \bigcap_{\forall (I, I^c, I^-)} \mathcal{L}_{S_{(I, I^c, I^-)}(x)}(x)^\circ.$$

⁶ (I, I^c, I^-) is partition of $\mathcal{I}_{GH}^{00}(z; t)$ means that $I \cup I^c \cup I^- = \mathcal{I}_{GH}^{00}(z; t)$ and $I \cap I^c = I \cap I^- = I^c \cap I^- = \emptyset$.

By [3, Theorem 3.2.2], we know that

$$\begin{aligned} \mathcal{L}_{S_{(I, I^c, I^-)}(x)}(x)^\circ &= \left\{ \sum_{i \in \mathcal{I}_g(x)} \eta_i^g \nabla g_i(x) + \sum_{i=1}^m \eta_i^h \nabla h_i(x) - \sum_{i \in \mathcal{I}_G(x; \bar{t})} \eta_i^G \nabla G_i(x) - \sum_{i \in \mathcal{I}_H(x; \bar{t})} \eta_i^H \nabla H_i(x) \right. \\ &\quad + \sum_{i \in \mathcal{I}_{GH}^{+0}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t)} \eta_i^\Phi \nabla \Phi_i^B(G(x), H(x); t) - \sum_{i \in I} \eta_i^G \nabla G_i(x) + \sum_{i \in I^c} \eta_i^G \nabla G_i(x) \\ &\quad + \sum_{i \in I} \eta_i^H \nabla H_i(x) - \sum_{i \in I^c} \eta_i^H \nabla H_i(x) + \sum_{i \in I^-} \delta_i^G \nabla G_i(x) + \sum_{i \in I^-} \delta_i^H \nabla H_i(x) : \\ &\quad \left. (\eta^g, \eta^G, \eta^H, \eta^\Phi) \geq 0 \right\}. \end{aligned}$$

For $v \in \mathcal{T}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ$, we have $v \in \mathcal{L}_{S_{(I, I^c, I^-)}(x)}(x)^\circ$ for any partition (I, I^c, I^-) of $\mathcal{I}_{GH}^{00}(x; t)$. If we fix I and set $I^- = \emptyset$, then there exists some multipliers η^h and $\eta^g, \eta^G, \eta^H, \eta^\Phi \geq 0$ so that

$$\begin{aligned} v &= \sum_{i \in \mathcal{I}_g(x)} \eta_i^g \nabla g_i(x) + \sum_{i=1}^m \eta_i^h \nabla h_i(x) - \sum_{i \in \mathcal{I}_G(x; \bar{t})} \eta_i^G \nabla G_i(x) - \sum_{i \in \mathcal{I}_H(x; \bar{t})} \eta_i^H \nabla H_i(x) \\ &\quad + \sum_{i \in \mathcal{I}_{GH}^{+0}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t)} \eta_i^\Phi \nabla \Phi_i^B(G(x), H(x); t) \\ &\quad - \sum_{i \in I} \eta_i^G \nabla G_i(x) + \sum_{i \in I^c} \eta_i^G \nabla G_i(x) - \sum_{i \in I} \eta_i^H \nabla H_i(x) + \sum_{i \in I^c} \eta_i^H \nabla H_i(x). \end{aligned}$$

Now, it also holds that $v \in \mathcal{L}_{S_{(I^c, I, I^-)}(x)}(x)^\circ$ and so there exists some multipliers η^h and $\eta^g, \eta^G, \eta^H, \eta^\Phi \geq 0$ such that

$$\begin{aligned} v &= \sum_{i \in \mathcal{I}_g(x)} \eta_i^g \nabla g_i(x) + \sum_{i=1}^m \eta_i^h \nabla h_i(x) - \sum_{i \in \mathcal{I}_G(x; \bar{t})} \eta_i^G \nabla G_i(x) - \sum_{i \in \mathcal{I}_H(x; \bar{t})} \eta_i^H \nabla H_i(x) \\ &\quad + \sum_{i \in \mathcal{I}_{GH}^{+0}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t)} \eta_i^\Phi \nabla \Phi_i^B(G(x), H(x); t) \\ &\quad + \sum_{i \in I} \eta_i^G \nabla G_i(x) - \sum_{i \in I^c} \eta_i^G \nabla G_i(x) + \sum_{i \in I} \eta_i^H \nabla H_i(x) - \sum_{i \in I^c} \eta_i^H \nabla H_i(x). \end{aligned}$$

By the construction of 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 η_i^G and η_i^H with indices i in $I \cup I^c$ vanish.

Therefore, $v \in \mathcal{L}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ$ and as v has been chosen arbitrarily then $\mathcal{T}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ \subseteq \mathcal{L}_{\mathcal{X}_{t, \bar{t}}^B}(x)^\circ$, which concludes the proof. \square

This result is sharp as shown by the following example, since Abadie CQ does not hold.

Example 4.3. Consider the problem

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

At $x^* = (0, 0)^T$ it holds that $\nabla \Phi^B(G(x), H(x); t) = (0, 0)^T$ and so $\mathcal{L}_{\mathcal{X}_{t, \bar{t}}^B}(x^*) = \mathbb{R}^2$, which is obviously different from the tangent cone at x^* for $t_2 \theta'(0) < t_1$ and $\bar{t} > 0$.

The following example shows that we cannot have a similar result using MPCC-GMFCQ.

Example 4.4. Consider the set

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

MPCC-GMFCQ holds at $x^* = (0, 0)^T$, since the gradients are linearly dependent but only with coefficients $\lambda^G = -\lambda^H$ that does not satisfy the condition given in Definition 2.7.

Now, we can choose a sequence of points such that $x^k \rightarrow x^*$ and

$$F_2(x^k; t_k) = 0, -t_{2,k}\theta'_{t_{1,k}}(H(x^k)) \rightarrow -1.$$

Since $\nabla G(x^*) = \nabla H(x^*)$ it holds that $\nabla F_2(x^*; 0) = (0 \ 0)^T$ and so MFCQ does not hold for (8).

It is disappointing to require MPCC-LICQ to obtain only GCQ, but when \mathcal{I}_{GH}^{00} is empty we obtain the stronger LICQ.

Theorem 4.5. Let $x^* \in \mathcal{Z}$, satisfying MPCC-LICQ. Then, there exists $t^* > 0$ and a neighborhood $U(x^*)$ of x^* such that

$$\forall t \in (0, t^*]: x \in U(x^*) \cap \mathcal{X}_{t,\bar{t}}^B \text{ and } \mathcal{I}_{GH}^{00}(x; t) = \emptyset \implies \text{LICQ holds at } x \text{ for } R_{t,\bar{t}}^B.$$

Proof. Let $x \in U(x^*) \cap \mathcal{X}_{t,\bar{t}}^B$ and t sufficiently small. We prove that the gradients of the constraints involved in $R_{t,\bar{t}}^B$ are linearly independent, by verifying that the trivial solution is the only solution to the following equation

$$\begin{aligned} 0 = & \sum_{i \in \mathcal{I}_g(x)} \eta_i^g \nabla g_i(x) + \sum_{i=1}^m \eta_i^h \nabla h_i(x) + \sum_{i \in \mathcal{I}_G(x;\bar{t})} \nabla G_i(x) \eta_i^G + \sum_{i \in \mathcal{I}_H(x;\bar{t})} \nabla H_i(x) \eta_i^H \\ & + \sum_{i \in \mathcal{I}_{GH}^{+0}(x;t)} \nabla G_i(x) (\eta_i^\Phi(F_{1i}(x;t) - F_{2i}(x;t)t_2\theta'_{t_1}(G_i(x)))) \\ & + \sum_{i \in \mathcal{I}_{GH}^{0+}(x;t)} \nabla H_i(x) (\eta_i^\Phi(F_{2i}(x;t) - F_{1i}(x;t)t_2\theta'_{t_1}(H_i(x)))) . \end{aligned}$$

MPCC-LICQ and the inclusions (9) gives that the only solution is the trivial one. \square

4.3 Convergence of the epsilon-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, can alter significantly the convergence analysis of relaxation methods as shown in [17, 20, 21] and [26].

Previous results in [21] prove convergence to C-stationary point for the relaxation from Scheel and Scholtes [30] and the one from Lin and Fukushima, [23], under some hypotheses on the sequence ϵ_k , respectively $\epsilon_k = O(t_k)$ and $\epsilon_k = o(t_k^2)$. Furthermore, the authors in [21] also provide a counter-example with a sequence converging to a W-stationary point if these conditions do not hold. Additionally, the authors in [21], prove that relaxations (KDB) and (KS) converge only to a W-stationary point and they require more hypotheses on the sequences ϵ_k and x_k to prove the convergence to a C- or a M-stationary point.

In the same way as in Theorem 4.1, we consider through this section a sequence of multipliers that should verify the stationary conditions. We denote for all $i \in \{1, \dots, q\}$

$$\begin{aligned} \lambda_i^{G,k} &:= \begin{cases} \eta_i^{G,k} + \eta_i^{\Phi,k} \left(t_{2,k}\theta'_{t_{1,k}}(G_i(x^k))F_{2i}(x^k; t_k) - F_{1i}(x^k; t_k) \right), & \text{if } F_{1i}(x^k; t_k) \geq -F_{2i}(x^k; t_k) \\ \eta_i^{G,k} + \eta_i^{\Phi,k} \left(F_{2i}(x^k; t_k) - t_{2,k}\theta'_{t_{1,k}}(G_i(x^k))F_{1i}(x^k; t_k) \right), & \text{if } F_{1i}(x^k; t_k) < -F_{2i}(x^k; t_k), \end{cases} \\ \lambda_i^{H,k} &:= \begin{cases} \eta_i^{H,k} + \eta_i^{\Phi,k} \left(t_{2,k}\theta'_{t_{1,k}}(H_i(x^k))F_{1i}(x^k; t_k) - F_{2i}(x^k; t_k) \right), & \text{if } F_{1i}(x^k; t_k) \geq -F_{2i}(x^k; t_k) \\ \eta_i^{H,k} + \eta_i^{\Phi,k} \left(F_{1i}(x^k; t_k) - t_{2,k}\theta'_{t_{1,k}}(H_i(x^k))F_{2i}(x^k; t_k) \right), & \text{if } F_{1i}(x^k; t_k) < -F_{2i}(x^k; t_k). \end{cases} \end{aligned} \tag{10}$$

In order to prove our main convergence theorem, we first prove a technical lemma.

Lemma 4.6. Consider the same assumptions as Theorem 4.7 below. Additionally, assume that for $i \in \mathcal{I}^{+0} \cup \mathcal{I}^{00}$, $\lim_{k \rightarrow \infty} \eta_i^{G,k} = \lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{1i}(x^k; t_k) = 0$ and for k sufficiently large $F_{1i}(x^k; t_k) \geq -F_{2i}(x^k; t_k)$. Then,

$$\lim_{k \rightarrow \infty} \frac{|\eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) F_{2i}(x^k; t_k)|}{\|\lambda^k\|_\infty} = 0.$$

As a consequence, $\lambda_i^{G,*} = 0$.

Proof. Without loss of generality, let us assume that $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) \neq 0$, otherwise we are done. $\|\lambda^k\|_\infty \geq |\lambda_i^{H,k}|$, so by (10) and $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{1i}(x^k; t_k) = 0$ it is sufficient to show that

$$\lim_{k \rightarrow \infty} \frac{|\eta_i^{\Phi,k} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k)) F_{2i}(x^k; t_k)|}{|\eta_i^{H,k} - \eta_i^{\Phi,k} F_{2i}(x^k; t_k)|} = 0. \quad (11)$$

We now consider two cases: either $\lim_{k \rightarrow \infty} \eta_i^{H,k} = 0$ or $\lim_{k \rightarrow \infty} \eta_i^{H,k} \neq 0$.

- If $\lim_{k \rightarrow \infty} \eta_i^{H,k} = 0$. Then, the left-hand side in (11) is equal to $\lim_{k \rightarrow \infty} t_{2,k} \theta'_{t_{1,k}}(G_i(x^k))$, which goes to zero by Lemma 3.4 as $t_{2,k} = o(t_{1,k})$.
- Consider the case, $\lim_{k \rightarrow \infty} \eta_i^{H,k} \neq 0$. Dividing by \bar{t}_k in the complementarity condition in (16) implies $H_i(x^k) \sim -\bar{t}_k$ as $\epsilon_k = o(\bar{t}_k)$. Thus, $H_i(x^k) < 0$ for k sufficiently large.

We prove that $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = 0$. Dividing by $H_i(x^k)$ in the complementarity condition in (17) gives for $H_i(x^k) \sim -\bar{t}_k$ that

$$\left| \eta_i^{\Phi,k} F_{2i}(x^k; t_k) \left(1 - \frac{t_{2,k} \theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)} \right) \right| \leq \frac{\epsilon_k}{|H_i(x^k)|} \rightarrow 0, \quad (12)$$

as $\epsilon_k = o(\bar{t}_k)$. However, $\lim_{k \rightarrow \infty} \frac{t_{2,k} \theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)} \neq 1$, otherwise $G_i(x^k) \leq 0$ and $|G_i(x^k)| \geq |H_i(x^k)|$ would yields to $\lim_{k \rightarrow \infty} \frac{t_{2,k} \theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)} \leq \lim_{k \rightarrow \infty} \frac{t_{2,k} \theta_{t_{1,k}}(H_i(x^k))}{H_i(x^k)} = 0$ as θ is non-decreasing and $t_{2,k} = o(t_{1,k})$. Therefore, (12) implies $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = 0$, which proves (11).

This completes the proof as $\lambda_i^{G,*} = 0$ by (10). \square

The following result proves convergence of the butterfly relaxation in this context.

Theorem 4.7. Given the three sequences $\{t_k\}, \{\bar{t}_k\}, \{\epsilon_k\}$ decreasing to zero and satisfying (4). Assume that $\epsilon_k = o(\max(|G(x^k)|, |H(x^k)|))$, $\epsilon_k = o(\bar{t}_k)$ and $t_{2,k} = o(t_{1,k})$. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ be a sequence of points that are ϵ_k -stationary points of R_{t_k, \bar{t}_k}^B for all $k \in \mathbb{N}$ with $x^k \rightarrow x^*$ such that MPCC-GCRSC holds at x^* . Then, x^* is a M-stationary point.

The notation $\epsilon_k = o(\max(|G(x^k)|, |H(x^k)|))$ means here that for all $i = 1, \dots, q$, $\epsilon_k = o(\max(|G_i(x^k)|, |H_i(x^k)|))$. For two sequences $\{g_k\}, \{h_k\}$ with the same signs for k sufficiently large, we also denote $g_k \sim h_k$ whenever $\lim_{k \rightarrow \infty} g_k/h_k = 1$.

Proof. The aim of the proof is to show that some sequence $\{\lambda^k\}$ verifies the assumptions of Theorem 2.11. The conclusion will follow by a direct application of this theorem.

The representation of $\nabla \Phi_i^B(G(x^k), H(x^k); t_k)$ immediately gives $\nabla \Phi_i^B(G(x^k), H(x^k); t_k) = 0, \forall i \in \mathcal{I}_{GH}^{00}(x^k; t_k)$ for all $k \in \mathbb{N}$. Thus, x^k being a ϵ_k stationary point for R_{t_k, \bar{t}_k}^B yields to

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

with $(\lambda^{g,k}, \lambda^{h,k}) = (\eta^{g,k}, \eta^{h,k})$ and $\lambda^{G,k}, \lambda^{H,k}$ defined in (10), and

$$|h_i(x^k)| \leq \epsilon_k \quad (i = 1, \dots, m), \quad (13)$$

$$g_i(x^k) \leq \epsilon_k, \quad \eta_i^{g,k} \geq 0, \quad \left| \eta_i^{g,k} g_i(x^k) \right| \leq \epsilon_k \quad (i = 1, \dots, p), \quad (14)$$

$$G_i(x^k) + \bar{t}_k \geq -\epsilon_k, \quad \eta_i^{G,k} \geq 0, \quad \left| \eta_i^{G,k} (G_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad (i = 1, \dots, q), \quad (15)$$

$$H_i(x^k) + \bar{t}_k \geq -\epsilon_k, \quad \eta_i^{H,k} \geq 0, \quad \left| \eta_i^{H,k} (H_i(x^k) + \bar{t}_k) \right| \leq \epsilon_k \quad (i = 1, \dots, q), \quad (16)$$

$$\Phi_i^B(G(x^k), H(x^k); t_k) \leq \epsilon_k, \quad \eta_i^{\Phi,k} \geq 0, \quad \left| \eta_i^{\Phi,k} \Phi_i^B(G(x^k), H(x^k); t_k) \right| \leq \epsilon_k \quad (i = 1, \dots, q). \quad (17)$$

Let us now check the three conditions of Theorem 2.11:

- (i) Obviously, we have $x^* \in \mathcal{Z}$ and $\nabla_x \mathcal{L}_{MPCC}(x^*, \lambda^h, \lambda^g, \lambda^G, \lambda^H) = 0$ since $\epsilon_k \downarrow 0$. Thus, the sequence $\{\lambda^k\}$ verifies assumption (i) of Theorem 2.11.
- (ii) The sequence $\{\lambda^k\}/\|\lambda^k\|_\infty$ is bounded and converges, up to a subsequence, to some limit denoted λ^* . We can assume without loss of generality that $\lim_{k \rightarrow \infty} \|\lambda^k\|_\infty \neq 0$, otherwise the result would be automatically true. Using that $\text{supp}(\lambda^g) \subset \text{supp}(\lambda^{g,k})$, (14) implies that $\lambda_i^{g,*} = 0$ for $i \notin \mathcal{I}_g(x^*)$.

Let us now show that for indices $i \in \mathcal{I}^{+0}$, $\lambda_i^{G,*} = 0$. The opposite case for indices $i \in \mathcal{I}^{0+}$ would follow in a completely similar way. So, let i be in \mathcal{I}^{+0} .

The complementarity condition in (15) gives that $\lim_{k \rightarrow \infty} \eta_i^{G,k} = 0$, since $\epsilon_k \downarrow 0$ and $G_i(x^k) + \bar{t}_k \rightarrow G_i(x^*) > 0$.

Note that we are necessarily in the case $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \geq 0$, as $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \rightarrow G_i(x^*) > 0$ ⁷. In this case we get $\lim_{k \rightarrow \infty} F_{1i}(x^k; t_k) \eta_i^{\Phi,k} = 0$ since $\left| \eta_i^{\Phi,k} \Phi_i^B(G(x^k), H(x^k); t_k) \right| \leq \epsilon_k$ by (17) and $\lim_{k \rightarrow \infty} F_{2i}(x^k; t_k) > 0$.

Since $\lim_{k \rightarrow \infty} F_{1i}(x^k; t_k) \eta_i^{\Phi,k} = \lim_{k \rightarrow \infty} \eta_i^{G,k} = 0$, applying Lemma 4.6, we obtain that $\lambda_i^{G,*} = 0$ for $i \in \mathcal{I}^{+0}$. Therefore, by symmetry, the sequence $\{\lambda^k\}$ satisfies assumption (ii) of Theorem 2.11.

- (iii) We now consider indices $i \in \mathcal{I}^{00}$. Our aim here is to verify assumption (3) (i.e. $\lambda_i^{G,*}, \lambda_i^{H,*}$ have M-stationary signs) in Theorem 2.11 for $\epsilon_k = o(\max(|G(x^k)|, |H(x^k)|))$.

Without loss of generality suppose that $\max(|G_i(x^k)|, |H_i(x^k)|) = |G_i(x^k)|$, and so $\lim_{k \rightarrow \infty} \frac{\epsilon_k}{|G_i(x^k)|} = 0$.

Let α (possibly infinite) be such that

$$\alpha := \lim_{k \rightarrow \infty} \frac{|G_i(x^k)|}{|t_{2,k} \theta_{t_{1,k}}(H_i(x^k))|}. \quad (18)$$

It should be noticed that $\alpha > 1$, otherwise for k sufficiently large there would exist a constant C such that $|G_i(x^k)| \leq C |t_{2,k} \theta_{t_{1,k}}(H_i(x^k))|$, which is a contradiction with $|G_i(x^k)| \geq |H_i(x^k)|$ and $t_{2,k} = o(t_{1,k})$.

Another consequence is that $F_{2i}(x^k; t_k) \sim G_i(x^k)$, since $F_{2i}(x^k; t_k) \leq G_i(x^k) + t_{2,k} \theta_{t_{1,k}}(|G_i(x^k)|)$ and by definition of functions θ s.

We consider separately the two cases $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \geq 0$ and $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) < 0$.

⁷We remind that $F_{1i}(x^k; t_k) = H_i(x) - t_{2,k} \theta_{t_{1,k}}(G_i(x))$ and $F_{2i}(x^k; t_k) = G_i(x) - t_{2,k} \theta_{t_{1,k}}(H_i(x))$. Thus, $\lim_{k \rightarrow \infty} (F_{2i}(x^k; t_k), F_{1i}(x^k; t_k)) = (G_i(x^*), 0)$ and $G_i(x^*) > 0$.

- a) When $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \geq 0$, dividing by $|G_i(x^k)|$ in the complementarity condition in (17) yields to

$$\frac{\epsilon_k}{|G_i(x^k)|} \geq \left| \eta_i^{\Phi,k} F_{1i}(x^k; t_k) \left(1 - \frac{t_{2,k} \theta_{t_{1,k}}(H_i(x^k))}{G_i(x^k)} \right) \right|,$$

so, $\eta_i^{\Phi,k} F_{1i}(x^k; t_k) \rightarrow 0$ since $\alpha > 1$.

Now, consider two cases either $\{\eta_i^{G,k}\}$ tends to zero or not. In the former case, the conclusion of case a) would follow by applying Lemma 4.6.

So, let $\lim_{k \rightarrow \infty} \lambda_i^{G,k} \neq 0$. Dividing by $G_i(x^k)$ in the complementarity condition in (15) gives $|\eta_i^{G,k}(1 + \bar{t}_k/G_i(x^k))| \leq \epsilon_k/|G_i(x^k)|$ and so $G_i(x^k) \sim -\bar{t}_k$.

Besides, it can be noted that for k sufficiently large there is no constant $C > 0$ such that $H_i(x^k) \leq C\epsilon_k$ as this would lead to a contradiction with $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \geq 0$. Indeed, as $H_i(x^k) \geq G_i(x^k)$, we would obtain

$$F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) \leq G_i(x^k) + C\epsilon_k - 2t_{2,k} \theta_{t_{1,k}}(G_i(x^k)),$$

which is negative for k sufficiently large by definition of θ and $\epsilon_k = o(G_i(x^k))$. So, $\epsilon_k = o(H_i(x^k))$ and $H_i(x^k) > 0$. Thus, dividing by $H_i(x^k)$ in the complementarity condition (17), we obtain $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = 0$. This concludes the case a), since $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = \lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{1i}(x^k; t_k) = 0$ gives that $(\lambda_i^{G,*}, \lambda_i^{H,*}) = \lim_{k \rightarrow \infty} (\eta_i^{G,k}, \eta_i^{H,k}) \geq 0$.

- b) When $F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k) < 0$, the complementarity condition in (17) gives $\left| \eta_i^{\Phi,k} F_{2i}(x^k; t_k) \right|^2 \leq 2\epsilon_k$, and dividing by $|G_i(x^k)|$ yields to

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

This implies that $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = 0$, by assumption on ϵ_k and $\alpha > 1$. Now, by definition of functions θ s and the triangle inequality, we get

$$|F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k)| \leq 2|G_i(x^k)| + 2t_{2,k} \theta_{t_{1,k}}(|G_i(x^k)|) \sim 2|G_i(x^k)|. \quad (19)$$

Using that $F_{2i}(x^k; t_k) \sim G_i(x^k)$ as noticed in the beginning of case (iii), we obtain that $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} F_{2i}(x^k; t_k) = \lim_{k \rightarrow \infty} \eta_i^{\Phi,k} G_i(x^k) = 0$. So, multiplying by $\eta_i^{\Phi,k}$ and going to the limit in (19) yields to $\lim_{k \rightarrow \infty} \eta_i^{\Phi,k} (F_{1i}(x^k; t_k) + F_{2i}(x^k; t_k)) = 0$. As a consequence, it holds that $(\lambda_i^{G,*}, \lambda_i^{H,*}) = \lim_{k \rightarrow \infty} (\eta_i^{G,k}, \eta_i^{H,k}) \geq 0$.

All in all, we completed cases a) and b), so the sequence $\{\lambda^k\}$ verifies assumption (3) of Theorem 2.11.

Since, MPCC-GCRSC holds at x^* , then Theorem 2.11 gives that x^* is a M-stationary point. \square

The assumption in Theorem 4.7 is not entirely satisfactory, since the sequence of parameter ϵ_k depends on the iterates. However, this is in the same vein than the existing results in [21]. Further research may try to exploit this weak point to propose more adequate conditions.

The following example, from [19], shows that the butterfly relaxation with $t_{2,k} = o(t_{1,k})$ may converge to an undesirable A-stationary point without the hypothesis that $\epsilon_k = o(\max(|G(x^k)|, |H(x^k)|))$.

Example 4.5. Consider the problem

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

Let $t_{2,k} = t_{1,k}^2$ and choose any positive sequences $\{t_{1,k}\}$ and $\{\epsilon_k\}$ such that $t_{1,k}, \epsilon_k \rightarrow 0$. Consider the following ϵ -stationary sequence

$$x^k = (\epsilon_k, \epsilon_k/2)^T, \quad \eta^{G,k} = 0, \quad \eta^{H,k} = 1 - \eta^{\Phi,k}(t_{1,k}^2 \theta_{t_{1,k}} \left(\frac{\epsilon_k}{2} \right) F_1(x^k; t_k) - F_2(x^k; t_k))$$

and

$$\eta^{\Phi,k} = \frac{1}{t_{1,k}^2 \theta_{t_{1,k}}(\epsilon_k) F_2(x^k; t_k) - F_1(x^k; t_k)}.$$

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

The ϵ -feasible set of the butterfly relaxation is similar to the relaxation from [30]. Therefore, it is not surprising that we can only expect to converge to a C-stationary point without strong hypothesis. Those issues clearly deserve a specific study that is left here for further research.

5 Numerical Results

In this section, we focus on the numerical implementation of the butterfly relaxation. Our aim is to compare the new method with the existing ones in the literature and to show some of its features. This comparison uses the collection of test problems MacMPEC [22]. This collection has been widely used in the literature to compare relaxation methods as in [17, 33, 14]. 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.

5.1 On the Implementation of the Butterfly Relaxation

Practical implementation could consider a slightly different model, by skipping the relaxation of the positivity constraint and adding a new parameter t_3 in order to shift the intersection of both wings to the point $(G(x), H(x)) = (t_3, t_3)$. This can be done by redefining $F_1(x; t_1, t_2, t_3)$ and $F_2(x; t_1, t_2, t_3)$ such that

$$\begin{aligned} F_1(x; t_1, t_2, t_3) &= H_i(x) - t_3 - t_2 \theta_{t_1}(G_i(x) - t_3), \\ F_2(x; t_1, t_2, t_3) &= G_i(x) - t_3 - t_2 \theta_{t_1}(H_i(x) - t_3). \end{aligned}$$

Even if we did not give any theoretical proof regarding this modified system, this modification does not alter the behavior of the butterfly relaxation. This formulation is clearly an extension of the relaxation (KS).

The numerical comparison of the butterfly relaxation with other existing methods considers three schemes:

1. $B_{(t_2=t_1)}$: $t_3 = 0, t_2 = t_1$;
2. $B_{(t_2=t_1^{3/2})}$: $t_3 = 0, t_2 = t_1^{3/2}$;
3. $B_{(t_3=t_2, 2t_2=t_1)}$: $t_3 = t_2, 2t_2 = t_1$.

In all these tests, we fixed $\bar{t} = 0$. Our tests concern many variants, not all of which covered by our analysis, but they give a broader insight of the new relaxations.

5.2 Comparison of the Relaxation Methods

We provide in this section and in Algorithm 1 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 the evaluation error of the objective function or the constraints.

Algorithm 1 is coded in Matlab and uses the AMPL API. R_{t_k} denotes the relaxed non-linear program associated with a generic relaxation, where except the butterfly methods the parameter $t_{1,k}$ does not play any role. At each step we compute x^{k+1} as a solution of R_{t_k} starting from x^k . Therefore, at each step the initial point is more likely to be infeasible for R_{t_k} . The iterative process stops when $t_{2,k}$ and $t_{1,k}$ are smaller than some tolerance, denoted p_{\min} which is set as 10^{-15} here, or when the solution x^{k+1} of R_{t_k} is considered an ϵ -solution of (MPCC). To consider x^{k+1} as a 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:

$$\nu_c(x) := \max(\|g(x) \circ \eta^g\|_\infty, \|h(x) \circ \eta^h\|_\infty, \|G(x) \circ \eta^G\|_\infty, \|H(x) \circ \eta^H\|_\infty, \|\Phi^B(x) \circ \eta^\Phi\|_\infty)$$
.

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

$$optimal(x) := \max(\nu_f(x), \nu_{comp}(x), \nu_c(x)).$$

A fourth criterion could be the dual feasibility, that is the norm of the gradient of the Lagrangian. However, solvers like SNOPT or MINOS do not use this criterion as a stopping criterion. One reason among other to discard such a criterion could be numerical issues implied by the degeneracy in the KKT conditions. In the case of an infeasible or unbounded sub-problem R_{t_k} , the algorithm stops and returns a certificate.

Data:

starting vector x^0 ; initial relaxation parameter t_0 ; update parameter $(\sigma_{t_1}, \sigma_{t_2}) \in (0, 1)^2$; p_{\min} the minimum parameter value; ϵ the precision tolerance ;

```

1 Begin :
2 Set  $k := 0$  ;
3 while  $\max(t_{2,k}, t_{1,k}) > p_{\min}$  and  $optimal(x^k) > \epsilon$  do
4   |  $x^{k+1}$  solution of  $R_{t_{1,k}, t_{2,k}}$  with  $x^k$  initial point;
5   |  $(t_{1,k+1}, t_{2,k+1}) := (t_{1,k}\sigma_{t_1}, t_{2,k}\sigma_{t_2})$  ;
6 return:  $f_{opt}$  the optimal value at the solution  $x_{opt}$  or a decision of infeasibility or unboundedness.
```

Algorithm 1: Basic Relaxation methods for (MPCC), with a relaxed non-linear program R_{t_k} .

Step 4 in Algorithm 1 is performed using three different solvers accessible through AMPL [10], that are SNOPT 7.2-8 [11], MINOS 5.51 [25] and IPOPT 3.12.4 [34] with their default parameters. A previous similar comparison in the literature in [14] only considered SNOPT to solve the sub-problems. We compare the butterfly schemes with the most popular relaxations SS from [29] and (KS). Moreover, we also take into account results of the non-linear programming solver without specific MPCC tuning denoted by 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 [14] we consider the value of the "intersection between G and H", which is (t, t) for (KS) and (Bu.), (\sqrt{t}, \sqrt{t}) for SS. 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. In [14], the authors consider as a stopping criterion the feasibility of the last non-linear

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

parametric program in particular by considering the complementarity constraint by the minimum component-wise. Table 2 provides our result with this criterion. We report elementary statistics by considering the percentage of success for each set of parameters. A problem is considered solved in this case if criteria (a) and (b) are satisfied.

Solver SNOPT	NL	SS	KS	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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 2: Sensitivity analysis for MacMPEC test problems considering the feasibility of (MPCC). Results are a percentage of success. best: percentage of success with the best set of parameters (independent of the problem), worst: percentage of success with the worst set of parameters, average: average percentage of success among the distribution of (T, s) , std: standard deviation.

First, we see that the method NL is giving decent results. It is not a surprise as was pointed out in [9]. 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 problems solved by NL. By using SNOPT, KS and butterfly with $t_2 = t_1^{3/2}$ methods get 1% of improvement and with IPOPT the method butterfly with $t_2 = t_1^{3/2}$ is the only one that attains 100%. The relaxation methods seem to make 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 $t_3 = t_2, 2t_2 = t_1$ is very competitive.

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 3. Once again we provide percentages of success.

In comparison with Table 2, this new criterion appears to be more selective. Independently of the solver, the relaxation methods with some correct choices of parameters provide improved results. Using SNOPT as a solver, the methods KS and butterfly gives the highest number of results. The method butterfly with $t_2 = t_1^{3/2}$ even improved the number of problems 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.

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.

Solver SNOPT	NL	SS	KS	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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	$B_{(t_2=t_1)}$	$B_{(t_3=t_2, 2t_2=t_1)}$	$B_{(t_2=t_1^{3/2})}$
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 3: Sensitivity analysis for MacMPEC test problems considering the optimality of (MPCC). Results are percentages of success. best: percentage of success with the best set of parameters, worst: percentage of success with the worst set of parameters, average: average percentage of success among the distribution of (T, s) , std: standard deviation.

Example 5.1. Consider the problem

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

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

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.

$x_1^0 \setminus x_2^0$	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 4: Sensitivity analysis depending on the initial point $(x_1^0, x_2^0, 0, 0)$ on Example 5.1 by using the butterfly relaxation method $t_2 = t_1^{3/2}$ 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 $x^2 + y^2 = 10$.

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 1 with $T = 0.5$ and $s = 0.01$. Table 4 shows that the butterfly relaxation with $t_2 = t_1^{3/2}$ 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 disturbing results are explained by Theorem 4.7 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 2.6 MPCC-GCQ does not hold at these points. Moreover, this example does not contradict the Theorem 4.7 since in particular MPCC-GCRSC is not verified at any feasible point of the problem.

6 Concluding Remarks

This paper proposes a new family of relaxation schemes for the mathematical program with complementarity constraints. We prove convergence of the method in the general case and shows that a specific relation between the parameters allows the method to converge to the desired M-stationary point. However, in the specific case where MPCC-LICQ, S-stationary conditions can be expected to hold at a local minimum. We prove that in the affine case the butterfly relaxation method converges to such a point without second order condition or conditions on the sequence x^k , which is a clear improvement over other methods.

Through the convergence study, we introduce the new and weak MPCC-GCRSC condition. We prove under this condition that some sequences of approximate stationary points are bounded.

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.

Future research will focus on the main difficulty regarding relaxation schemes that are the convergence of approximate stationary sequences. A discussion regarding the former problem has been initiated in [21] and appeal further discussion.

Funding

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

References

- [1] 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.
- [2] 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.
- [3] Mokhtar S. Bazaraa and Chitharanjan Marakada Shetty. *Foundations of optimization*, volume 122. Springer Science & Business Media, 2012.
- [4] 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.
- [5] Jean-Pierre Dussault, Mounir Haddou, Abdeslam Kadrani and Tangi Migot. How to compute the local minimum of the MPCC. *Optimization Online*, 2017.
- [6] 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.
- [7] 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.
- [8] Michael L. Flegel and Christian Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints*. Springer, 2006.
- [9] Roger Fletcher and Sven Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004.
- [10] Robert Fourer, David Gay, and Brian Kernighan. *Ampl*, volume 119. Boyd & Fraser, 1993.
- [11] Philip E. Gill, Walter Murray, and Michael A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005.
- [12] 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.
- [13] Mounir Haddou. A new class of smoothing methods for mathematical programs with equilibrium constraints. *Pacific Journal of Optimization*, 5(1), 2009.
- [14] 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, 2013.
- [15] 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.
- [16] 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.
- [17] 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.

- [18] Christian Kanzow and Alexandra Schwartz. Mathematical programs with equilibrium constraints: enhanced fritz john-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] Sven Leyffer. Macmpec: Ampl collection of mpecs. *Argonne National Laboratory. Available at www.mcs.anl.gov/leyffer/MacMPEC*, 2000.
- [23] 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.
- [24] Tangi Migot. *Contributions to numerical methods for complementarity problems and mathematical programs with complementarity constraints*. PhD thesis, IRMAR, INSA de Rennes, 2017.
- [25] Bruce A. Murtagh and Michael A. Saunders. Minos 5.0 user’s guide. Technical report, DTIC Document, 1983.
- [26] Alberto Ramos. Mathematical Programms with Equilibrium Constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization Online*, 2016.
- [27] Alberto Ramos. Two new weak constraint qualifications for mathematical programs with equilibrium constraints and applications. *Optimization Online*, 2017.
- [28] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- [29] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.
- [30] Stefan Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.
- [31] 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.
- [32] 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.
- [33] Sonja Steffensen and Michael Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.
- [34] 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.

A Proof of a Technical Lemma

In the proof of Theorem 4.4 and Theorem 4.5 we use the following lemma that links the gradients of G and H with the gradients of $F_1(x; t)$ and $F_2(x; t)$.

Lemma A.1. *Let (I, I^c, I^-) be any partition of $\mathcal{I}_{GH}^{00}(x; t)$. Assume that the gradients*

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

are linearly independent. Then, LICQ holds at x for (8).

Proof. We show that the gradients of the constraints of (8) are positively linearly independent. For this purpose, we prove that the trivial solution is the only solution to the equation

$$\begin{aligned} 0 = & \sum_{i \in \mathcal{I}_g(x)} \eta_i^g \nabla g_i(x) + \sum_{i=1}^m \eta_i^h \nabla h_i(x) - \sum_{i \in \mathcal{I}_G(x; \bar{t})} \eta_i^G \nabla G_i(x) - \sum_{i \in \mathcal{I}_H(x; \bar{t})} \eta_i^H \nabla H_i(x) \\ & + \sum_{i \in \mathcal{I}_{GH}^{+0}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t)} \eta_i^\Phi \nabla \Phi_i^B(G(x), H(x); t) \\ & + \sum_{i \in \mathcal{I}_{GH}^{00}(x; t)} \left(\nu_i^{F_1(x; t)} - \mu_i^{F_1(x; t)} + \delta_i^{F_1(x; t)} \right) \nabla F_{1i}(x; t) + \left(-\nu_i^{F_2(x; t)} + \mu_i^{F_2(x; t)} + \delta_i^{F_2(x; t)} \right) \nabla F_{2i}(x; t), \end{aligned}$$

where $\text{supp}(\eta^g) \subset \mathcal{I}_g(x)$, $\text{supp}(\eta^G) \subset \mathcal{I}_G(x; \bar{t})$, $\text{supp}(\eta^H) \subset \mathcal{I}_H(x; \bar{t})$, $\text{supp}(\eta^\Phi) \subset \mathcal{I}_{GH}^{+0}(x; t) \cup \mathcal{I}_{GH}^{0+}(x; t)$, $\text{supp}(\nu^{F_1(x; t)}) \subset I$, $\text{supp}(\nu^{F_2(x; t)}) \subset I$, $\text{supp}(\mu^{F_1(x; t)}) \subset I^c$, $\text{supp}(\mu^{F_2(x; t)}) \subset I^c$, $\text{supp}(\delta^{F_1(x; t)}) \subset I^-$, $\text{supp}(\delta^{F_2(x; t)}) \subset I^-$ where $I \cup I^c \cup I^- = \mathcal{I}_{GH}^{00}(x; t)$ and I, I^c, I^- have two by two empty intersection.

By definition of $F_1(x; t)$ and $F_2(x; t)$ it holds that

$$\begin{aligned} \nabla F_{1i}(x; t) &= \nabla H_i(x) - t_2 \theta'_{t_1}(G_i(x)) \nabla G_i(x), \\ \nabla F_{2i}(x; t) &= \nabla G_i(x) - t_2 \theta'_{t_1}(H_i(x)) \nabla H_i(x). \end{aligned}$$

The gradient of $\Phi^B(G(x), H(x); t)$ is given by Lemma 3.3.

We now replace those gradients in the equation above

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

with

$$\begin{aligned} \lambda_i^G &= -\eta_i^G + \eta_i^\Phi F_{1i}(x; t) - \left(\eta_i^\Phi F_{2i}(x; t) + \nu_i^{F_1(x; t)} - \mu_i^{F_1(x; t)} + \delta_i^{F_1(x; t)} \right) t_2 \theta'_{t_1}(G_i(x)) \\ &\quad - \nu_i^{F_2(x; t)} + \mu_i^{F_2(x; t)} + \delta_i^{F_2(x; t)}, \\ \lambda_i^H &= -\eta_i^H + \eta_i^\Phi F_{2i}(x; t) - \left(\eta_i^\Phi F_{1i}(x; t) - \nu_i^{F_2(x; t)} + \mu_i^{F_2(x; t)} + \delta_i^{F_2(x; t)} \right) t_2 \theta'_{t_1}(H_i(x)) \\ &\quad + \nu_i^{F_1(x; t)} - \mu_i^{F_1(x; t)} + \delta_i^{F_1(x; t)}. \end{aligned}$$

By linear independence assumption, we obtain

$$\begin{aligned} \eta^g &= 0, \eta^h = 0, \eta^G = 0, \eta^H = 0, \eta_i^\Phi = 0 \ \forall i \in \mathcal{I}_{GH}^{0+}(x; t) \cup \mathcal{I}_{GH}^{+0}(x; t), \\ -\nu_i^{F_1(x; t)} t_2 \theta'_{t_1}(G_i(x)) - \nu_i^{F_2(x; t)} &= 0 \text{ and } \nu_i^{F_2(x; t)} t_2 \theta'_{t_1}(H_i(x)) + \nu_i^{F_1(x; t)} = 0 \ \forall i \in I, \\ \mu_i^{F_1(x; t)} t_2 \theta'_{t_1}(G_i(x)) + \mu_i^{F_2(x; t)} &= 0 \text{ and } -\mu_i^{F_2(x; t)} t_2 \theta'_{t_1}(H_i(x)) - \mu_i^{F_1(x; t)} = 0 \ \forall i \in I^c \\ -\delta_i^{F_1(x; t)} t_2 \theta'_{t_1}(G_i(x)) + \delta_i^{F_2(x; t)} &= 0 \text{ and } -\delta_i^{F_2(x; t)} t_2 \theta'_{t_1}(H_i(x)) + \delta_i^{F_1(x; t)} = 0 \ \forall i \in I^-. \end{aligned}$$

So, it follows for $i \in I^-$ that

$$\delta_i^{F_2(x;t)} = \delta_i^{F_1(x;t)} t_2 \theta'_{t_1}(G_i(x)) \text{ and } \delta_i^{F_1(x;t)} = \delta_i^{F_2(x;t)} t_2 \theta'_{t_1}(H_i(x)).$$

So $\delta_i^{F_1(x;t)} = \delta_i^{F_2(x;t)} = 0$, since $i \in \mathcal{I}_{GH}^{00}(x;t)$ gives

$$t_2 \theta'_{t_1}(G_i(x)) t_2 \theta'_{t_1}(H_i(x)) = t_2 \theta'_{t_1}(0) t_2 \theta'_{t_1}(0) < 1$$

by properties of θ and (4). Similarly, we get $\mu_i^{F_1(x;t)} = \mu_i^{F_2(x;t)} = \nu_i^{F_2(x;t)} = \nu_i^{F_1(x;t)} = 0$. \square

B Useful Lemmas

Lemma B.1. [32, Lemma 7.1] 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 $\text{supp}(\alpha^*) \subseteq \text{supp}(\alpha)$, $\text{supp}(\beta^*) \subseteq \text{supp}(\beta)$ and

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

such that the vectors

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

are linearly independent.

In our proofs we use the following results from [32] to compute the tangent cone of $\mathcal{X}_{t,\bar{t}}^B$ and its polar.

Lemma B.2. [Lemma 8.10, [32]] For all $t > 0$ and all x feasible for $R_{t,\bar{t}}^B$,

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

where $S_{(I,I^c,I^-)}(x)$ is defined in equation 8.