

# The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints

Jean-Pierre Dussault, Mounir Haddou, Tangi Migot

▶ **To cite this version:**

Jean-Pierre Dussault, Mounir Haddou, Tangi Migot. The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints. 2016. hal-01525399v5

**HAL Id: hal-01525399**
**https://hal.science/hal-01525399v5**

Preprint submitted on 23 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The New Butterfly Relaxation Method for Mathematical Programs with Complementarity Constraints

J.-P. Dussault and M. Haddou and T. Migot

**Abstract** We propose a new family of relaxation schemes for mathematical programs with complementarity constraints. 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 an M-stationary point. A stronger convergence result is also proved in the affine case. A comprehensive numerical comparison between existing relaxations methods is performed on the library of test problems MacMPEC which shows promising results for our new method.

Preprint version of an article published in `https://link.springer.com/chapter/10.1007/978-981-16-1819-2_3`

## 1 Introduction

We consider the Mathematical Program with Complementarity Constraints (MPCC)

$$\min_{x\in\mathbb{R}^n} \; f(x) \text{ s.t. } g(x) \leq 0, h(x) = 0, 0 \leq G(x) \perp H(x) \geq 0, \tag{1}$$

J.-P. Dussault
Département d'Informatique, faculté des Sciences, Université de Sherbrooke, Canada, e-mail: `Jean-Pierre.Dussault@USherbrooke.ca`

M. Haddou
Univ Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France, e-mail: `Mounir.Haddou@insa-rennes.fr`

T. Migot
Département d'Informatique, faculté des Sciences, Université de Sherbrooke, Canada, e-mail: `tangi.migot@gmail.com`

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$ and $G, H : \mathbb{R}^n \to \mathbb{R}^q$ 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 $u \geq 0, v \geq 0$ and $u^T v = 0$ for all $i \in \{1, \ldots, q\}$. 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.

The MPCC is a non-linear program, but with a special structure since, apart from the usual equality and inequality constraints, they have the additional complementarity constraints, which may be equivalently rewritten as

$$G_i(x) \geq 0, H_i(x) \geq 0, G_i(x)H_i(x) \leq 0, \quad \forall i \in \{1, \ldots, q\}. \tag{2}$$

A popular approach to tackle a non-linear program 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 don't hold in general for (1) due to the complementarity constraint. For instance, the classical Mangasarian-Fromowitz constraint qualification is violated at any feasible point [38].

During the past two decades, many researchers introduced necessary optimality conditions such as the Clarke (C-), Mordukhovich (M-), strong (S-), and Bouligand (B-) stationarity conditions for the MPCC; see, e.g., [11, 18, 30, 35, 36, 37, 38]. Among these stationarities, the B-stationarity is known to be a good candidate for optimality, but since it is computationally difficult, it is rarely used in algorithmic analysis; the S-stationarity is the strongest and equivalent to the KKT conditions (see, e.g., [10, 15]), but its interest is reduced since it does not always hold for the MPCC. The M-stationarity, which has already widely been investigated (see, e.g.,[11, 18, 26, 35, 36, 37, 38]), is the most relevant concept since it is the weakest necessary condition holding, under suitable constraint qualifications, at any local minimizer of the MPCC and is computationaly tractable.

The feasible set of the MPCC involves a complementarity constraint equivalent to $G(x) = 0$ OR $H(x) = 0$. This is a *thin* set exhibiting some irregularity when $G(x) = 0$ AND $H(x) = 0$. It is this thinness that makes constraint qualifications fail at any feasible point. 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 (1) is obtained similarly to a homotopy technique.

These methods have been suggested in the literature back in 2000 by Scheel and Scholtes in [30, 31]. Their natural approach was later extended by Demiguel, Friedlander, Nogales and Scholtes in [6] In [23], Lin and Fukushima improved this relaxation by expressing the same set with two constraints instead of three. This improvement leads to an 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 a 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 points, called C-stationary points, may still happen, the convergence to M-stationary points being guaranteed only under some second-order condition.

A new perspective for those schemes has been given in [17] relaxing the constraints (2) by $t \geq 0$ as

$$(G_i(x) - t)(H_i(x) - t) \leq 0, \quad \forall i \in \{1, \ldots, q\}. \tag{3}$$

This approximation scheme converges as $t$ decreases to 0 under classical assumptions to M-stationary points without second-order or strict complementarity-type conditions. This is not a relaxation since the feasible domain of (1) is not included in the feasible set of the sub-problems. The method has been extended as a relaxation method in [19] through an NCP function $\phi$:

$$\phi(G_i(x) - t, H_i(x) - t) \leq 0, \quad \forall i \in \{1, \ldots, q\}, \tag{4}$$

where $\phi(a, b) := \{ab, \text{ if } a + b \geq 0, -\frac{1}{2}(a^2 + b^2), \text{ otherwise}\}$.

The main aim of this paper is to continue this discussion and extend the relaxation of Kanzow and Schwartz [19] by introducing the new butterfly relaxation:

$$\phi\left(G_i(x) - t_2\theta_{t_1}(H_i(x)), H_i(x) - t_2\theta_{t_1}(G_i(x))\right) \leq 0, \quad \forall i \in \{1, \ldots, q\}.$$

This new method handling two relaxing parameters, $t_1$ and $t_2$, allows a non-linear perturbation, $\theta$, of the domain. Thus, we extend the butterfly relaxation introduced in [8] for mathematical program with vanishing constraints to the case of complementarity constraints.

The following example shows that the butterfly relaxation may improve the relaxations from [17] and [19]. Indeed, it illustrates an example where their is no sequence of stationary point[1] that converges to a non-optimal point.

*Example 1*

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

In this example, there are two stationary points: an S-stationary point $(1, 0)$ that is the global minimum and an M-stationary point $(0, 0)$, which is not a local minimum.[2] Unlike the relaxations (3) and (4) where for $t_k = \frac{1}{k}$ a sequence $x^k = (t_k, \ 2t_k)^T$, with the Lagrange multiplier associated to the regularized constraint $\lambda^{\Phi,k} = k$, may converge to $(0, 0)$ as $k$ goes to infinity, there is no sequence of stationary points that converges to this undesirable point with the butterfly relaxation.

Our main contributions in this paper are the following:

---

[1] Definitions of stationary points of a non-linear program at the beginning of Section 2.1 page 4.

[2] Definitions of M- and S-stationarity points are given in Definition 3 page 6.

1. We prove convergence of the butterfly relaxation scheme to A-stationary points, and to M-stationary points for $t_{2,k} = o(t_{1,k})$.
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.
3. We prove that the butterfly relaxation scheme computing approximate stationary points at each step converges to an 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 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 literature. In Section 3, we define the relaxation scheme with the new butterfly relaxation. In Section 4, we prove theoretical results on convergence and the existence of multipliers of the relaxed sub-problems. We also provide an analysis of the convergence of approximate stationary points. We also generalize the situation of Example 1 to illustrate a situation where the non-linear perturbation allows us to escape from undesirable points. In Section 5, we provide an extensive numerical study by giving details on the implementation and a comparison with other methods. Finally, in Section 6, we discuss some perspectives of this work.

## 2 Preliminaries

In this section, we introduce classical notations and definitions for non-linear programs and mathematical programs with complementarity constraints used in the sequel.

### 2.1 Non-Linear Programming

Let a general non-linear program be

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) \leq 0,\ h(x) = 0, \tag{5}$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$. Denote $\mathcal{F}$ the feasible region of (5), and $\mathcal{I}_g(x) := \{i \in \{1, ..., m\} \ : \ g_i(x) = 0\}$ the set of active indices. The Lagrangian function of (5) is defined as $\mathcal{L}(x, \lambda) := f(x) + g(x)^T \lambda^g + h(x)^T \lambda^h$, where $\lambda = (\lambda^g, \lambda^h) \in \mathbb{R}^m \times \mathbb{R}^p$ is the vector of Lagrange multipliers.

We call a KKT point a couple $(x, \lambda)$ 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 call $x$ a stationary point if there exists $\lambda$ such that $(x, \lambda)$ is a KKT point. We remind that the tangent cone of a set $X \subseteq \mathbb{R}^n$ at $x^* \in X$ is a closed cone defined by

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

Another useful tool for our study is the linearized cone of (5) at $x^* \in \mathcal{F}$ defined by

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

In the context of solving non-linear programs, that is finding a local minimum of (5), 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 (5) that satisfies a constraint qualification, then there exists a Lagrange multiplier $\lambda^*$ such that $(x^*, \lambda^*)$ is a KKT point of (5). 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 1** Let $x^* \in \mathcal{F}$.

(a) Linear Independence CQ (LICQ) holds at $x^*$ if the family of gradients $\{\nabla g_i(x^*)\ (i \in \mathcal{I}_g(x^*)),\ \nabla h_i(x^*)\ (i = 1, ..., p)\}$ is linearly independent.
(b) Mangasarian-Fromovitz CQ (MFCQ) holds at $x^*$ if the family of gradients $\{\nabla h_i(x^*)\ (i = 1, \ldots, p)\}$ 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 $\nabla h_i(x^*)^T d = 0\ (i = 1, \ldots, p)$.

*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, so that under MFCQ, the only solution of $\sum_{i\in\mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^p \lambda_i^h \nabla h_i(x^*) = 0$ with $\lambda_i^g \geq 0, \forall i \in \mathcal{I}_g(x^*)$ is the trivial solution.

A local minimum is characterized by the fact that there is no feasible descent direction for the objective function of (5), that is $-\nabla f(x^*) \in \mathcal{T}_{\mathcal{F}}(x^*)^\circ$, where $\mathcal{T}^\circ$ denotes the polar cone of $\mathcal{T}$. Given a cone $K \subseteq \mathbb{R}^n$, the polar of $K$ is the cone defined by $K^\circ := \{z \in \mathbb{R}^n | z^T x \leq 0, \forall x \in K\}$. On the other hand, the KKT conditions build $\nabla f$ 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 the context of numerical computations, it is almost never possible to compute stationary points. Hence, it is of interest to consider $\epsilon$-stationary points.

**Definition 2** Given a general non-linear program (5) and $\epsilon \geq 0$. We say that $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{m+p}$ is an $\epsilon$-KKT point if it satisfies

$$\|\nabla_x \mathcal{L}(x, \lambda)\|_\infty \leq \epsilon, \|h(x)\|_\infty \leq \epsilon,$$
$$g_i(x) \leq \epsilon,\ \lambda_i^g \geq 0,\ \left|\lambda_i^g g_i(x)\right| \leq \epsilon,\ \forall i \in \{1, \ldots, m\}.$$

We say that $x$ is an $\epsilon$-stationary point if there exists $\lambda$ such that $(x, \lambda)$ is an $\epsilon$-KKT point.

## 2.2 Mathematical Programs with Complementarity Constraints

We now specialize the general notions above to our specific case of (1). Let $\mathcal{Z}$ be the set of feasible points of (1). 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 (1) defined as

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

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

We introduce more stationary concepts as in [24, 26, 27, 30, 35, 36, 37]. Those concepts are needed for two reasons:

- unless assuming a restrictive constraint qualification, a local minimizer $x^*$ may fail to be a stationary point, so that optimality conditions need to be weakened in order to obtain a necessary condition;
- when analyzing cluster points of algorithms, other weak stationarity conditions appear naturally.

**Definition 3** A point $x^* \in \mathcal{Z}$ is said

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

$$\begin{aligned}
&\nabla_x \mathcal{L}_{MPCC}(x^*, \lambda) = 0,\\
&\lambda^g \geq 0, \lambda^g_i = 0, \forall i \notin \mathcal{I}_g,\\
&\lambda^G_i = 0, \forall i \in \mathcal{I}^{+0}, \text{ and, } \lambda^H_i = 0, \forall i \in \mathcal{I}^{0+};
\end{aligned}$$

- C-stationary, if it is W-stationary and $\lambda^G_i \lambda^H_i \geq 0, \forall i \in \mathcal{I}^{00}$;
- A-stationary, if it is W-stationary and $\lambda^G_i \geq 0$ or $\lambda^H_i \geq 0, \forall i \in \mathcal{I}^{00}$;
- M-stationary, if it is W-stationary and either $\lambda^G_i > 0, \lambda^H_i > 0$ or $\lambda^G_i \lambda^H_i = 0$, $\forall i \in \mathcal{I}^{00}$;
- S-stationary, if it is W-stationary and $\lambda^G_i \geq 0,\ \lambda^H_i \geq 0, \forall i \in \mathcal{I}^{00}$.

Relations between these definitions are straightforward from the definitions.

As pointed out in [10], strong stationarity is equivalent to the standard KKT conditions of an MPCC. In order to guarantee that a local minimum $x^*$ of (1) is a stationary point in any of the previous senses, one needs to assume that a constraint qualification (CQ) is satisfied in $x^*$. Since most standard CQs are violated at any feasible point of (1), many MPCC-analogues of these CQs have been developed. Here, we mention only those needed later.

**Definition 4** Let $x^* \in \mathcal{Z}$.

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

$$\sum_{i \in \mathcal{I}_g(x^*)} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) - \sum_{i \in \mathcal{I}^{0+} \cup \mathcal{I}^{00}} \lambda_i^G \nabla G_i(x^*) - \sum_{i \in \mathcal{I}^{+0} \cup \mathcal{I}^{00}} \lambda_i^H \nabla H_i(x^*) = 0 \quad (6)$$

   is the trivial solution.
2. MPCC-MFCQ holds at $x^*$ if the only solution of (6) with $\lambda_i^g \geq 0, \forall i \in \mathcal{I}_g(x^*)$, is the trivial solution.
3. MPCC-GMFCQ holds at $x^*$ if the only solution of (6) with $\lambda_i^g \geq 0, \forall i \in \mathcal{I}_g(x^*)$, and either $\lambda_i^G \lambda_i^H = 0$ or $\lambda_i^G > 0, \lambda_i^H > 0, \forall i \in \mathcal{I}^{00}$ 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. Note that each of these CQs implies that a local minimum is M-stationary, see [9, 36], but only MPCC-LICQ is sufficient to guarantee strong stationarity of a local minimum, see [10, 24, 28]. The MPCC-LICQ is among the first MPCC-tailored constraint qualifications and may already be found in [24] and [30], the MPCC-MFCQ was introduced in [30] and presented in the form above in [16].

## 3 The Butterfly Relaxation Method

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

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

Then, for $t_1 > 0$, we introduce $\theta_{t_1}(x) := \theta\left(\frac{x}{t_1}\right)$ if $x \geq 0$, and completed in a smooth way for negative values by considering $\theta_{t_1}(x) = x\theta'(0)/t_1$ if $x < 0$.

*Example 2* Examples of such functions are

$$\theta^1_{t_1}(x) := \{\frac{x}{x + t_1}, \text{ for } x \geq 0, \quad \frac{x}{t_1}, \text{ for } x < 0.\},$$

and

$$\theta^2_{t_1}(x) := \{1 - \exp^{-\frac{x}{t_1}}, \text{ for } x \geq 0, \quad \frac{x}{t_1}, \text{ for } x < 0.\}.$$

Those functions have already been used in the context of complementarity constraints, for instance, in [1, 2].

To simplify the notation, we denote $t := (t_1, t_2)$. Using this family of functions, we denote

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

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

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

This new relaxation uses two parameters $t_1$ and $t_2$ chosen such that

$$t_2\theta'(0) < t_1. \tag{8}$$

This condition ensures that the intersection of 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.

The parametric non-linear program related to the butterfly relaxation of the complementarity constraints defined in (7), and augmented with a regularization of the non-negativity constraints parametrized by $\bar{t}$, is given by

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t } x \in \mathcal{X}^B_{t,\bar{t}}, \tag{$R^B_{t,\bar{t}}$}$$

with

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

where $e$ denotes the vector of all ones.

This method is similar to the methods (3) from [17] and (4) 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. A comparison of the feasible set of these relaxations can be seen in Figure 1.

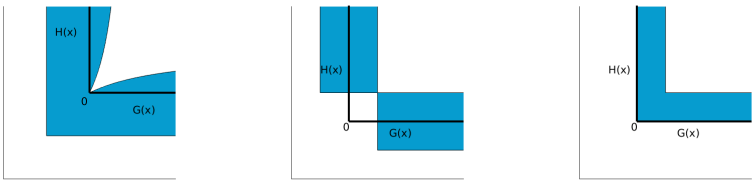



Fig. 1: The feasible set of the butterfly relaxation, the approximation from [17] and the relaxation from [19].

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

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

Additionally, by definition of the relaxation mapping, it holds

$$\Phi_i^B(G(x),H(x);t) = 0 \Longrightarrow F_{1i}(x;t) + F_{2i}(x;t) \geq 0.$$

The following two lemmas give more insights on the relaxation.

**Lemma 1** *Let $x \in \mathcal{X}_{t,\bar{t}}^B$, then it is true for the relaxation* (7) *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)* $i \in \mathcal{I}_{GH}(x;t) \Longrightarrow G_i(x) \geq 0, H_i(x) \geq 0.$

***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, i.e. $G_i(x) < 0$, it would follow that

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

which would be negative using (8). Note that previous inequality holds true since, by definition of the function $\theta$, it holds that $t_2\theta_{t_1}(z) = t_2\theta'(0)z/t_1$ for $z \leq 0$. Finally, $G_i(x) \geq 0$ implies that $H_i(x) \geq 0$ since $F_{1i}(x;t) = 0$. □

The following lemma state two of the key features of the relaxation and follows from the observation that $\Phi_i^B(G(x),H(x);t) \to G_i(x)H_i(x)$ as $\|t\| \downarrow 0$.

**Lemma 2** *The set $\mathcal{X}_{t,\bar{t}}^B$ satisfy the following properties:*

1. $\lim\limits_{\|t,\bar{t}\|\to 0} \mathcal{X}_{t,\bar{t}}^B = \mathcal{Z}$ *where the limit is assumed pointwise*[3]*;*

[3] $\lim\limits_{k\to\infty} A^k = A$ pointwise means that for all sequences $\{x^k\}$ with $x^k \in A^k$ for all $k$ implies $\lim\limits_{k\to\infty} x^k \in A$ and for any $x^* \in A$ there exists a sequence $x^k$ with $x^k \in A^k$ such that $\lim\limits_{k\to\infty} x^k = x^*$.

*2.* $\bigcap_{t,\bar{t}\geq 0} \mathcal{X}^B_{t,\bar{t}} = \mathcal{Z}$.

If the feasible set of the (1) is non-empty, then the feasible sets of the relaxed subproblems are also non-empty for all $t \geq 0$. If for some parameter $t \geq 0$ the set $\mathcal{X}^B_{t,\bar{t}}$ is empty, then it implies that $\mathcal{Z}$ is empty. Finally, if a local minimum of $R^B_{t,\bar{t}}$ already belongs to $\mathcal{Z}$, then it is a local minimum of the (1).

Finally, numerical results will be presented in Sect. 5 and will 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 two technical lemmas. Direct computation gives the gradient of $\Phi^B(G(x), H(x); t)$ in the following lemma.

**Lemma 3** *For all $i \in \{1, \ldots, q\}$, the gradient of $\Phi^B_i(G(x), H(x); t)$ w.r.t. $x$ for the relaxation* (7) *is given by*

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

The following lemma illustrates the behavior of functions $\theta_{t_1}$ and their derivatives.

**Lemma 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*

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

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

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

***Proof*** First part of the lemma follows from the definition of functions $\theta_{t_{1,k}}$. Indeed, it holds for all $x \in \mathbb{R}_+$ that $\theta_{t_{1,k}}(x) \in [0, 1]$. Therefore, $\lim_{k\to\infty} t_{2,k}\theta_{t_{1,k}}(x_k) = 0$.

Second part of the lemma uses the fact that functions $\theta_{t_{1,k}}$ are defined as perspective functions, that is for all $z \in \mathbb{R}_+$

$$\theta_{t_{1,k}}(z) = \theta\left(\frac{z}{r_k}\right),$$

and so, computing the derivative gives

$$t_{2,k}\theta'_{t_{1,k}}(z) = \frac{t_{2,k}}{t_{1,k}}\theta'\left(\frac{z}{t_{1,k}}\right).$$

So, for $k$ sufficiently large $t_{1,k} \leq z$, and by concavity of $\theta_r$, we get

$$0 \leq \lim_{k\to\infty} t_{2,k}\theta'_{t_{1,k}}(z) \leq \lim_{k\to\infty} t_{2,k}\theta'_{t_{1,k}}(t_{1,k}) = \lim_{k\to\infty} \frac{t_{1,k}}{t_{1,k}}\theta'(1),$$

and the result follows. □

We focus on the sequel on the properties of these new relaxation schemes. We prove that the method converges to an A-stationary point in Theorem 1 and to an M-stationary point, Theorem 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 3).

The main motivation to consider relaxation methods for (1) 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 Theoretical Properties

The study of the theoretical properties of the butterfly relaxation method is split into three parts: convergence of the sequence of stationary points, the 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 an M-stationary point or, at least, provide a certificate if we converge to an undesirable point.

We prove in Theorem 1 that the butterfly relaxation converges to an A-stationary point. This result is improved to convergence to M-stationary points for some choices on the parameters $t_2$ and $t_1$ in Theorem 2.

Finally, we prove convergence to S-stationary point in a specific situation in Theorem 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 1** *Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying* (8) *and decreasing to zero. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ be a sequence of KKT*

*points of* $(R^B_{t,\bar{t}})$ *with* $x^k \to x^*$ *such that MPCC-MFCQ 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 3 through the stationary points of $(R^B_{t,\bar{t}})$.The representation of $\nabla\Phi^B$ immediately gives $\nabla\Phi^B_i(G(x^k), H(x^k); t_k) = 0,\ \forall i \in \mathcal{I}^{00}_{GH}(x^k; t_k)$. Thus, we can write

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

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

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

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

First, by (9), it holds that $\nabla\mathcal{L}_{MPCC}(x^k, \lambda^k) = 0$ for all $k$. Thus, the first condition of the W-stationary conditions is satisfied. Moreover, by definition of $\{\lambda^{g,k}\}$ it holds that $\mathcal{I}_g(x^k) \subseteq \mathcal{I}_g(x^*)$ and so $\lim_{k\to\infty} \lambda_i^{g,k} = 0, \forall i \notin \mathcal{I}_g(x^*)$.

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 (9) and since by (8) 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 that [4]

$$\|\lambda^k\|_\infty = \|\eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k} \circ F_2(x^k; t_k), \eta^{\Phi,k} \circ F_1(x^k; t_k))\|_\infty, \tag{10}$$

where $\circ$ denotes the componentwise product of two vectors.

We now verify that: $\lambda_i^{G,k} / \|\lambda^k\|_\infty \to 0$ for indices $i \in \mathcal{I}^{+0}$. By symmetry, it would follow that $\lambda_i^{H,k} / \|\lambda^k\|_\infty \to 0$ for indices $i \in \mathcal{I}^{0+}$.

Let $i \in \mathcal{I}^{+0}$. Clearly $i \in \mathcal{I}^{0+}_{GH}(x^k; t_k)$ as otherwise $G_i(x^k) = \bar{t}_k$ for $i \in \mathcal{I}_G(x^k; t_k)$ or $G_i(x^k) = t_{2,k}\theta_{t_{1,k}}(H_i(x^k))$ for $i \in \mathcal{I}^{+0}_{GH}(x^k; t_k)$ which in both cases, for $k$ sufficiently large, contradicts the fact $G_i(x^k) \to G_i(x^*) > 0$. Now, $i \in \mathcal{I}^{0+}_{GH}(x^k; t_k)$

[4] For indices $i \in \mathcal{I}^{0+}_{GH}(x^k; t_k)$ (symmetry for indices $i \in \mathcal{I}^{+0}_{GH}(x^k; t_k)$), 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.

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, $\|\lambda^k\|_\infty \geq |\eta_i^{\Phi,k} F_{2i}(x^k; t_k)|$ by (10), thus

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

since $G_i(x^k) \to G(x^*) > 0$ and using Lemma 4.

Now, let us prove that the sequence $\{\lambda^k\}$ is bounded. Assume by contradiction that is not bounded, then the sequence $\{\lambda^k / \|\lambda^k\|_\infty\}$ is bounded and converges, up to a subsequence to a non-trivial limit $\hat{\lambda}$. Therefore, dividing (9) by $\|\lambda^k\|_\infty$ and passing to the limit gives

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

which leads to a contradiction since $x^*$ satisfies MPCC-MFCQ.

So, the sequence $\{\lambda^k\}$ is bounded, hence $\lambda_i^{G,k} \to 0, \forall i \in \mathcal{I}^{+0}$ and $\lambda_i^{H,k} \to 0, \forall i \in \mathcal{I}^{0+}$. Therefore, $x^*$ is a W-stationary point of the MPCC.

Finally, let us now verify that $x^*$ is an A-stationary point. Denote $\lambda^*$ the limit, up to a subsequence, of the sequence $\{\lambda^k\}$. 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. □

The following example shows that the result of Theorem 1 is sharp since convergence cannot be ensured, assuming only that MPCC-GMFCQ holds at the limit point.

*Example 3* 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 $(0,0)^T$. 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 $x^k \to (0,0)^T$ 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 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 an M-stationary point.

**Theorem 2** *Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying* (8) *and decreasing to zero. Let $\{x^k, \eta^{g,k}, \eta^{h,k}, \eta^{G,k}, \eta^{H,k}, \eta^{\Phi,k}\}$ be a sequence of KKT points of $(R^B_{t,\bar{t}})$ with $x^k \to x^*$ such that MPCC-GMFCQ holds at $x^*$. If $t_{2,k} = o(t_{1,k})$, then, $x^*$ is an M-stationary point.*

***Proof*** In the proof of Theorem 1 we already showed that: $\nabla \mathcal{L}_{MPCC}(x^k, \lambda^k) = 0$ for all $k$, $\lim_{k\to\infty} \lambda_i^{g,k} = 0, \forall i \notin \mathcal{I}_g(x^*)$, $\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty = 0, \forall i \in \mathcal{I}^{+0}$, and $\lim_{k\to\infty} \lambda_i^{H,k}/\|\lambda^k\|_\infty = 0, \forall i \in \mathcal{I}^{0+}$.

Let us now check that either $\lim_{k\to\infty} \lambda_i^{G,k}\lambda_i^{H,k}/\|\lambda^k\|_\infty^2 = 0$ or $\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty > 0$, $\lim_{k\to\infty} \lambda_i^{H,k}/\|\lambda^k\|_\infty > 0$ using the contrapositive, i.e.

$$\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty < 0 \Longrightarrow \lim_{k\to\infty} \lambda_i^{H,k}/\|\lambda^k\|_\infty = 0,$$

and the other case will be similar by symmetry.

Let $i \in \mathcal{I}^{00}$. $\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty < 0$ implies that $i \in \mathcal{I}^{+0}_{GH}(x^k; t_k)$ for $k$ sufficiently large by definition of $\lambda_i^{G,k}$ as the function $\theta$ is non-decreasing. 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, $\lim_{k\to\infty} t_{2,k}\theta'_{t_{1,k}}(H_i(x^k)) = 0$ by Lemma 4 with $t_{2,k} = o(t_{1,k})$ and

$$\lim_{k\to\infty} \eta_i^{\Phi,k} F_{1i}(x^k; t_k)/\|\lambda^k\|_\infty = \lim_{k\to\infty} -\lambda_i^{G,k}/\|\lambda^k\|_\infty < \infty.$$

Thus, $\lim_{k\to\infty} \lambda_i^{H,k}/\|\lambda^k\|_\infty = 0$.

Finally, following the same reasoning as in the proof of Theorem 1, using MPCC-GMFCQ, the sequence $\{\lambda^k\}$ is bounded, and $x^*$ is an M-stationary point. □

The following example shows that this result is sharp, since it illustrates a situation where MPCC-GMFCQ does not hold and the method converges to an undesirable W-stationary point. This phenomenon only happens if the sequence of multipliers defined in (9) is unbounded.

*Example 4* Consider the problem

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

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 $(\lambda^G, \lambda^H = 0)$.

MPCC-GMFCQ 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

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

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

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

The following result motivated by Example 1 shows that the butterfly relaxation may improve its behavior in some specific cases. Example 1 also indicates that this cannot be expected with the other relaxations. In the sequel, we denote $\text{supp}(z) := \{i \mid z_i \neq 0\}$ the non-zero indices of $z$.

**Theorem 3** *Assume that $f, g, h, G, H$ are affine functions. Given two sequences $\{t_k\}$ and $\{\bar{t}_k\}$ of positive parameters satisfying* (8) *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 KKT points of $(R^B_{t,\bar{t}})$ with $x^k \to x^*$ such that MPCC-LICQ holds at $x^*$. If $t_{2,k} = o(t_{1,k})$, and, for all $k$ sufficiently large*

$$\text{supp}(\eta^{\Phi,k}) \cap \left(\mathcal{I}^{+0} \cup \mathcal{I}^{0+}\right) = \emptyset, \tag{11}$$

*then, $x^*$ is an S-stationary point.*

***Proof*** Theorem 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 $(R^B_{t,\bar{t}})$.

We already mention in the proof of Theorem 1 that for all $k$ it holds

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

where we omit the dependence in $k$ in the expression of the gradients, since they are constant by linear/affine assumption. Clearly, for $k$ sufficiently large, it holds that $\text{supp}(\lambda^{g,k}) \subseteq \mathcal{I}_g(x^*)$, $\text{supp}(\lambda^{G,k}) \subset \mathcal{I}^{0+} \cup \mathcal{I}^{00}$ and $\text{supp}(\lambda^{H,k}) \subseteq \mathcal{I}^{00} \cup \mathcal{I}^{+0}$ by (11).

Now, by continuity, linear independence of these gradients holds in a neighbourhood of $x^*$. So, we get finite convergence of the $\lambda^k$, and for $k$ sufficiently large it holds

$$\lambda^{g,k} = \lambda^{g,\infty},\ \lambda^{h,k} = \lambda^{h,\infty},\ \lambda^{G,k} = \lambda^{G,\infty},\ \lambda^{H,k} = \lambda^{H,\infty}. \tag{12}$$

Let $i \in \mathcal{I}^{00} \cap \text{supp}(\eta^{\Phi,\infty})$, where we remind that $\text{supp}(\eta^{\Phi,\infty}) \subseteq \text{supp}(\eta^{\Phi,k}) \subseteq \mathcal{I}_{GH}(x^k;t_k)$. 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 $i \in \mathcal{I}^{+0}_{GH}$ by definition of the multipliers in (9) 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 proved in Theorem 2 that $\lambda^{G,k}$ and $\lambda^{H,k}$ have bounded limits, so by Lemma 4 with $t_{2,k} = o(t_{1,k})$ we have $\lim_{k\to\infty} \lambda_i^{H,k} = 0$. By (12), we get $\eta_i^{\Phi,k} = 0$ for all $k$ sufficiently large, which contradicts $\lambda_i^{G,\infty} < 0$. □

## 4.2 Existence of Lagrange Multipliers for the Relaxed Sub-Problems

In this section, we study some regularity properties of the relaxed non-linear programs. Indeed, 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.

**Theorem 4** *Let $x^* \in \mathcal{Z}$, satisfying MPCC-LICQ. Then, there exists $t^* > 0$ and a neighborhood $U(x^*)$ of $x^*$ such that:*

$$\forall t \in (0, t^*] : \ x \in U(x^*) \cap \mathcal{X}^B_{t,\bar{t}} \Longrightarrow \textit{standard GCQ holds at } x \textit{ for } (R^B_{t,\bar{t}}).$$

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

The linearized cone of$(R^B_{t,\bar{t}})$ is given by

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

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

Let us compute the polar of the tangent cone. Consider the following set of non-linear constraints parametrized by $z \in \mathcal{X}^B_{t,\bar{t}}$ and a partition $(I, I^c, I^-)$ of $\mathcal{I}^{00}_{GH}(z; t)$[5], defined by

$$\begin{aligned}\mathrm{S}_{(I,I^c,I^-)}(z) := \{x \in \mathbb{R}^n \mid & g(x) \le 0, h(x) = 0, G(x) \ge -\bar{t}e, \ H(x) \ge -\bar{t}e, \\ & \Phi_i^B(G(x), H(x); t) \le 0, \ i \notin \mathcal{I}^{00}_{GH}(z; t), \\ & F_{1i}(x; t) \le 0, \ F_{2i}(x; t) \ge 0, \ i \in I, \\ & F_{1i}(x; t) \ge 0, \ F_{2i}(x; t) \le 0, \ i \in I^c, \\ & F_{1i}(x; t) \le 0, \ F_{2i}(x; t) \le 0, \ i \in I^-\}.\end{aligned} \tag{13}$$

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

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

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

$$\begin{aligned} &\mathcal{I}_g(x) \subseteq \mathcal{I}_g(x^*),\ \mathcal{I}_G(x;\bar{t}) \subseteq \mathcal{I}^{00} \cup \mathcal{I}^{0+},\ \mathcal{I}_H(x;\bar{t}) \subseteq \mathcal{I}^{+0} \cup \mathcal{I}^{00}, \\ &\mathcal{I}^{00}_{GH}(x;t) \cup \mathcal{I}^{+0}_{GH}(x;t) \subseteq \mathcal{I}^{00} \cup \mathcal{I}^{0+}, \\ &\mathcal{I}^{00}_{GH}(x;t) \cup \mathcal{I}^{0+}_{GH}(x;t) \subseteq \mathcal{I}^{+0} \cup \mathcal{I}^{00}. \end{aligned} \tag{14}$$

Therefore, by Lemma 6, LICQ holds for (13) at $x$. Furthermore, by [32, Lemma 8.10], and since LICQ in particular implies Abadie CQ it follows that

$$\mathcal{T}_{\mathcal{X}^B_{t,\bar{t}}}(x) = \bigcup_{\forall (I,I^c,I^-)} \mathcal{T}_{\mathrm{S}_{(I,I^c,I^-)}(x)}(x) = \bigcup_{\forall (I,I^c,I^-)} \mathscr{L}_{\mathrm{S}_{(I,I^c,I^-)}(x)}(x).$$

By [5, Theorem 3.1.9], passing to the polar, we get

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

By [5, Theorem 3.2.2], we know that

$$\begin{aligned} \mathscr{L}_{\mathrm{S}_{(I,I^c,I^-)}(x)}(x)^\circ = \{ &\sum_{i \in \mathcal{I}_g(x)} \eta^g_i \nabla g_i(x) + \sum_{i=1}^{p} \eta^h_i \nabla h_i(x) \\ &- \sum_{i \in \mathcal{I}_G(x;\bar{t})} \eta^G_i \nabla G_i(x) - \sum_{i \in \mathcal{I}_H(x;\bar{t})} \eta^H_i \nabla H_i(x) \\ &+ \sum_{i \in \mathcal{I}^{+0}_{GH}(x;t) \cup \mathcal{I}^{0+}_{GH}(x;t)} \eta^\Phi_i \nabla \Phi^B_i(G(x),H(x);t) \\ &- \sum_{i \in I} \eta^G_i \nabla G_i(x) + \sum_{i \in I^c} \eta^G_i \nabla G_i(x) \\ &+ \sum_{i \in I} \eta^H_i \nabla H_i(x) - \sum_{i \in I^c} \eta^H_i \nabla H_i(x) \\ &+ \sum_{i \in I^-} \delta^G_i \nabla G_i(x) + \sum_{i \in I^-} \delta^H_i \nabla H_i(x) : (\eta^g,\eta^G,\eta^H,\eta^\Phi) \geq 0\}. \end{aligned}$$

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

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

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

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

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 $\eta_i^G$ and $\eta_i^H$ with indices $i$ in $I \cup I^c$ vanish.
Therefore, $v \in \mathscr{L}_{\mathcal{X}^B_{t,\bar{t}}}(x)^\circ$ and as $v$ has been chosen arbitrarily then $\mathcal{T}_{\mathcal{X}^B_{t,\bar{t}}}(x)^\circ \subseteq \mathscr{L}_{\mathcal{X}^B_{t,\bar{t}}}(x)^\circ$, which concludes the proof. □

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

*Example 5* 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 $\mathscr{L}_{\mathcal{X}^B_{t,\bar{t}}}(x^*) = \mathbb{R}^2$, 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 6* Consider the set

$$C := \{(x_1,x_2)^T \mid 0 \leq x_1 + x_2^2 \perp x_1 \geq 0\}.$$

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 4.

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

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

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 (13).

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

**Theorem 5** *Let $x^* \in \mathcal{Z}$, satisfying MPCC-LICQ. Then, there exists $t^* > 0$ and a neighborhood $U(x^*)$ of $x^*$ such that*

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

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

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

MPCC-LICQ and the inclusions (14) give that the only solution is the trivial one. □

### 4.3 Convergence of the epsilon-stationary points

Non-linear programming algorithms usually compute sequences of approximate stationary points or $\epsilon$-stationary points (see Definition 2). We present below in relations (16)-(21) our specific definition and hypothesis of $\epsilon$-stationary points. This approach, which has become an active subject recently, can significantly alter the convergence analysis of relaxation methods, as shown in [17, 20, 21] and [29].

Previous results in [21] prove convergence to C-stationary point for the relaxation from Scheel and Scholtes [31] and the one from Lin and Fukushima [23], under some hypotheses on the sequence $\epsilon_k$, 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 (3) and (4) converge only to a W-stationary point and they require more hypotheses on the sequences $\epsilon_k$ and $x_k$ to prove the convergence to a C- or an M-stationary point.

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

$\{1,\dots,q\}$

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

$$\lambda_i^{H,k} := \begin{cases} \eta_i^{H,k} + \eta_i^{\Phi,k}\left(t_{2,k}\theta'_{t_{1,k}}(H_i(x^k))F_{1i}(x^k;t_k) - F_{2i}(x^k;t_k)\right), \\ \text{if } F_{1i}(x^k;t_k) \geq -F_{2i}(x^k;t_k) \\ \eta_i^{H,k} + \eta_i^{\Phi,k}\left(F_{1i}(x^k;t_k) - t_{2,k}\theta'_{t_{1,k}}(H_i(x^k))F_{2i}(x^k;t_k)\right), \\ \text{if } F_{1i}(x^k;t_k) < -F_{2i}(x^k;t_k). \end{cases} \tag{15}$$

The representation of $\nabla\Phi_i^B(G(x^k),H(x^k);t_k)$ immediately gives for all $i \in \mathcal{I}_{GH}^{00}(x^k;t_k)$ and all $k$ that $\nabla\Phi_i^B(G(x^k),H(x^k);t_k) = 0$. Thus, $x^k$ being a $\epsilon_k$ stationary point for $(R_{t,\bar{t}}^B)$ satisfies

$$\left\|\mathcal{L}_{MPCC}(x^k,\lambda^k)\right\|_\infty \leq \epsilon_k,$$

with $(\lambda^{g,k},\lambda^{h,k}) = (\eta^{g,k},\eta^{h,k})$ and $\lambda^{G,k},\lambda^{H,k}$ defined in (15), and

$$|h_i(x^k)| \leq \epsilon_k, \forall i = 1,\dots,p, \tag{16}$$

$$g_i(x^k) \leq \epsilon_k,\ \eta_i^{g,k} \geq 0,\ \left|\eta_i^{g,k} g_i(x^k)\right| \leq \epsilon_k, \forall i = 1,\dots,m, \tag{17}$$

$$G_i(x^k) + \bar{t}_k \geq -\epsilon_k,\ \eta_i^{G,k} \geq 0,\ \left|\eta_i^{G,k}(G_i(x^k) + \bar{t}_k)\right| \leq \epsilon_k, \forall i = 1,\dots,q, \tag{18}$$

$$H_i(x^k) + \bar{t}_k \geq -\epsilon_k,\ \eta_i^{H,k} \geq 0,\ \left|\eta_i^{H,k}(H_i(x^k) + \bar{t}_k)\right| \leq \epsilon_k, \forall i = 1,\dots,q, \tag{19}$$

$$\Phi_i^B(G(x^k),H(x^k);t_k) \leq \epsilon_k, \eta_i^{\Phi,k} \geq 0, \tag{20}$$

$$\left|\eta_i^{\Phi,k}\Phi_i^B(G(x^k),H(x^k);t_k)\right| \leq \epsilon_k, \forall i = 1,\dots,q. \tag{21}$$

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

**Lemma 5** *Consider the same assumptions as Theorem 6 below. Additionally, assume that for* $i \in \mathcal{I}^{+0} \cup \mathcal{I}^{00}$, $\lim_{k\to\infty} \eta_i^{G,k} = \lim_{k\to\infty} \eta_i^{\Phi,k} F_{1i}(x^k;t_k) = 0$ *and for* $k$ *sufficiently large* $F_{1i}(x^k;t_k) \geq -F_{2i}(x^k;t_k)$. *Then,*

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

*As a consequence,* $\lim_{k\to\infty} \frac{\lambda_i^{G,k}}{\|\lambda^k\|} = 0$.

***Proof*** Without loss of generality, let us assume that $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{2i}(x^k;t_k) \neq 0$, otherwise we are done. Since $\|\lambda^k\|_\infty \geq |\lambda_i^{H,k}|$, by (15) and $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{1i}(x^k;t_k) =$

0, it is sufficient to show that

$$\lim_{k\to\infty} \frac{|\eta_i^{\Phi,k} t_{2,k}\theta'_{t_{1,k}}(G_i(x^k))F_{2i}(x^k;t_k)|}{|\eta_i^{H,k} - \eta_i^{\Phi,k}F_{2i}(x^k;t_k)|} = 0. \tag{22}$$

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

- If $\lim\limits_{k\to\infty}\eta_i^{H,k} = 0$. Then, the left-hand side in (22) is equal to $\lim\limits_{k\to\infty} t_{2,k}\theta'_{t_{1,k}}(G_i(x^k))$, which goes to zero by Lemma 4 as $t_{2,k} = o(t_{1,k})$.
- Consider the case, $\lim\limits_{k\to\infty}\eta_i^{H,k} \neq 0$. Dividing by $\bar{t}_k$ in the compelmentarity condition in (19) 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 $\lim\limits_{k\to\infty}\eta_i^{\Phi,k}F_{2i}(x^k;t_k) = 0$. Dividing by $H_i(x^k)$ in the complementarity condition in (20) gives for $H_i(x^k) \sim -\bar{t}_k$ that

$$\left|\eta_i^{\Phi,k}F_{2i}(x^k;t_k)\left(1 - \frac{t_{2,k}\theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)}\right)\right| \leq \frac{\epsilon_k}{|H_i(x^k)|} \to 0, \tag{23}$$

  as $\epsilon_k = o(\bar{t}_k)$. However, $\lim\limits_{k\to\infty}\frac{t_{2,k}\theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)} \neq 1$, otherwise $G_i(x^k) \leq 0$ and $|G_i(x^k)| \geq |H_i(x^k)|$ would yield $\lim\limits_{k\to\infty}\frac{t_{2,k}\theta_{t_{1,k}}(G_i(x^k))}{H_i(x^k)} \leq \lim\limits_{k\to\infty}\frac{t_{2,k}\theta_{t_{1,k}}(H_i(x^k))}{H_i(x^k)} = 0$ as $\theta$ is non-decreasing and $t_{2,k} = o(t_{1,k})$. Therefore, (22) follows as (23) implies $\lim\limits_{k\to\infty}\eta_i^{\Phi,k}F_{2i}(x^k;t_k) = 0$. □

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

**Theorem 6** *Given the three sequences $\{t_k\}, \{\bar{t}_k\}, \{\epsilon_k\}$ decreasing to zero and satisfying* (8)*. 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 $\epsilon_k$-KKT points of $(R^B_{t,\bar{t}})$ with $x^k \to x^*$ such that MPCC-GMFCQ holds at $x^*$. Then, $x^*$ is an M-stationary point.*

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

***Proof*** Proceeding in the same way as Theorem 2, we verify that:

(i) $x^* \in \mathcal{Z}$, $\lim\limits_{k\to\infty}\nabla\mathcal{L}_{MPCC}(x^k, \lambda^k) = 0$, $\lim\limits_{k\to\infty}\lambda_i^{g,k} = 0, \forall i \notin \mathcal{I}_g(x^*)$,

(ii) $\lim\limits_{k\to\infty}\lambda_i^{G,k}/\|\lambda^k\|_\infty = 0, \forall i \in \mathcal{I}^{+0}$, $\lim\limits_{k\to\infty}\lambda_i^{H,k}/\|\lambda^k\|_\infty = 0, \forall i \in \mathcal{I}^{0+}$,

(iii) $\lim\limits_{k\to\infty}\lambda_i^{G,k}\lambda_i^{H,k}/\|\lambda^k\|_\infty^2 = 0$ or $\lim\limits_{k\to\infty}\lambda_i^{G,k}/\|\lambda^k\|_\infty > 0$, $\lim\limits_{k\to\infty}\lambda_i^{H,k}/\|\lambda^k\|_\infty > 0, \forall i \in \mathcal{I}^{00}$.

Clearly (i) follows from the stationarity of $x^k$ as $\epsilon_k \downarrow 0$.

Let us now show that for indices $i \in \mathcal{I}^{+0}$, $\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty = 0$. 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 (18) gives that $\lim_{k\to\infty} \eta_i^{G,k} = 0$, since $\epsilon_k \downarrow 0$ and $G_i(x^k) + \bar{t}_k \to G_i(x^*) > 0$.

Note that we are necessarily in the case $F_{1i}(x^k;t_k) + F_{2i}(x^k;t_k) \geq 0$, as $F_{1i}(x^k;t_k)+F_{2i}(x^k;t_k) \to G_i(x^*) > 0$[6]. In this case we get $\lim_{k\to\infty} F_{1i}(x^k;t_k)\eta_i^{\Phi,k} = 0$ since $\left|\eta_i^{\Phi,k}\Phi_i^B(G(x^k),H(x^k);t_k)\right| \leq \epsilon_k$ by (20) and $\lim_{k\to\infty} F_{2i}(x^k;t_k) > 0$.

Since $\lim_{k\to\infty} F_{1i}(x^k;t_k)\eta_i^{\Phi,k} = \lim_{k\to\infty} \eta_i^{G,k} = 0$, applying Lemma 5, we obtain that $\lim_{k\to\infty} \lambda_i^{G,k}/\|\lambda^k\|_\infty = 0$ for $i \in \mathcal{I}^{+0}$.

We now consider indices $i \in \mathcal{I}^{00}$. Without loss of generality suppose that $\max(|G_i(x^k)|, |H_i(x^k)|) = |G_i(x^k)|$, and so $\lim_{k\to\infty} \frac{\epsilon_k}{|G_i(x^k)|} = 0$. Let $\alpha$ (possibly infinite) be such that

$$\alpha := \lim_{k\to\infty} \frac{|G_i(x^k)|}{|t_{2,k}\theta_{t_{1,k}}(H_i(x^k))|}. \tag{24}$$

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 $\theta$s.

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

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 (20) 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, $\eta_i^{\Phi,k} F_{1i}(x^k;t_k) \to 0$ 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 5.

So, let $\lim_{k\to\infty} \lambda_i^{G,k} \neq 0$. Dividing by $G_i(x^k)$ in the complementarity condition in (18) 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

[6] We 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, $\lim_{k\to\infty}(F_{2i}(x^k;t_k), F_{1i}(x^k;t_k)) = (G_i(x^*), 0)$ and $G_i(x^*) > 0$.

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

which is negative for $k$ sufficiently large by definition of $\theta$ 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 (20), we obtain $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{2i}(x^k;t_k) = 0$. This concludes the case a), since $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{2i}(x^k;t_k) = \lim_{k\to\infty} \eta_i^{\Phi,k} F_{1i}(x^k;t_k) = 0$ gives that $(\lambda_i^{G,*}, \lambda_i^{H,*}) = \lim_{k\to\infty} (\eta^{G,k}, \eta^{H,k}) \geq 0$.

b) When $F_{1i}(x^k;t_k) + F_{2i}(x^k;t_k) < 0$, the complementarity condition in (20) gives $\left|\eta_i^{\Phi,k} F_{2i}(x^k;t_k)^2\right| \leq 2\epsilon_k$, 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 $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{2i}(x^k;t_k) = 0$, by assumption on $\epsilon_k$ and $\alpha > 1$. Now, by definition of functions $\theta$s and the triangle inequality, we get

$$\left|F_{1i}(x^k;t_k) + F_{2i}(x^k;t_k)\right| \leq 2|G_i(x^k)| + 2t_{2,k}\theta_{t_{1,k}}(|G_i(x^k)|) \sim 2|G_i(x^k)|. \tag{25}$$

Using that $F_{2i}(x^k;t_k) \sim G_i(x^k)$ as noticed in the beginning of case (iii), we obtain that $\lim_{k\to\infty} \eta_i^{\Phi,k} F_{2i}(x^k;t_k) = \lim_{k\to\infty} \eta_i^{\Phi,k} G_i(x^k) = 0$. So, multiplying by $\eta_i^{\Phi,k}$ and going to the limit in (25) yields to $\lim_{k\to\infty} \eta_i^{\Phi,k}\left(F_{1i}(x^k;t_k) + F_{2i}(x^k;t_k)\right) = 0$. As a consequence, it holds that $\lim_{k\to\infty} (\lambda_i^{G,k}, \lambda_i^{H,k})/\|\lambda^k\|_\infty = \lim_{k\to\infty} (\eta^{G,k}, \eta^{H,k}) \geq 0$.

All in all, we completed cases a) and b), so (iii) is satisfied.

Finally, since (i)-(ii)-(iii) are satisfied, we conclude as in Theorem 2, so that under MPCC-GMFCQ the sequence $\{\lambda^k\}$ is bounded and $x^*$ is an M-stationary point. □

The assumption in Theorem 6 is not entirely satisfactory since the sequence of parameter $\epsilon_k$ depends on the iterates. However, this is in the same vein as the existing results in [7, 21]. Further research may try to exploit this weak point to propose more adequate conditions.

Another benefit of considering approximate stationary points is that they may exist even so the assumptions presented in previous section are not satisfied, see [3, 4].

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 7* Consider the problem

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

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

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

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 $\epsilon$-feasible set of the butterfly relaxation is similar to the relaxation from [31]. Therefore, it is not surprising that we can only expect to converge to a C-stationary point without strong hypotheses. Those issues clearly deserve a specific study that is left 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 [16, 17, 33]. The test problems included in MacMPEC are extracted from the literature and real-world applications.

### 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

$$\begin{aligned} F_{1i}(x; t_1, t_2, t_3) &= H_i(x) - t_3 - t_2\theta_{t_1}(G_i(x) - t_3), \\ F_{2i}(x; t_1, t_2, t_3) &= G_i(x) - t_3 - t_2\theta_{t_1}(H_i(x) - t_3). \end{aligned}$$

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 (4).

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 Algorithm 1 some more details on the implementation and the comparison between relaxation methods. It is to be noted that we aim to compare the methods and so no attempt to optimize any method has been carried out. We use 101 test problems from MacMPEC, which omit 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 $\epsilon$-solution of (1). To consider $x^{k+1}$ as a $\epsilon$-solution, with $\epsilon$ 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:

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

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 $\epsilon$. Using these criteria, we define a measure of optimality

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

A fourth criterion could be the dual feasibility, which is the norm of the gradient of the Lagrangian. However, solvers like SNOPT or MINOS do not use this criterion as a stopping criterion, but use the gradient of the Lagrangian scaled by the norm of the Lagrange multiplier. One reason among others 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; $\epsilon$ 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 $\quad x^{k+1}$ solution of $R_{t_{1,k},t_{2,k}}$ with $x^k$ initial point;
5 $\quad (t_{1k+1}, t_{2k+1}) := (t_{1,k}\sigma_{t_1}, t_{2,k}\sigma_{t_2})$ ;
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 (1), with a relaxed non-linear program $R_{t_k}$.

Step 4 in Algorithm 1 is performed using three different solvers accessible through AMPL [13], that are SNOPT 7.2-8 [14], MINOS 5.51 [25] and IPOPT 3.12.4 [34] with their default parameters. A previous similar comparison in the literature in [16] only considered SNOPT to solve the sub-problems. We compare the butterfly schemes with the most popular relaxations SS from [30] and (4). We also take into account the 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 [16] we consider the value of the "intersection between G and H", which is $(t, t)$ for (4) and (7), $(\sqrt{t}, \sqrt{t})$ 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 $\sigma_t$ as described in Table 1. In [16], the authors consider as a stopping criterion the feasibility of the

| Relaxation | NL | SS | KS | Butterfly |
|---|---|---|---|---|
| $t_0$ | none | $T^2$ | $T$ | $T$ |
| $\sigma_t$ | none | $S^2$ | $S$ | $S$ |

Table 1: Parameter links among the methods

last non-linear 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 if criteria (a) and (b) are satisfied.

First, we see that the method NL is giving decent results. It is not a surprise, as was pointed out in [12]. Practical implementation of relaxation methods would select the best choice of parameters so that we focus most of our attention on the line 'best'. In

| **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 (1). 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.

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 (1) 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 give the highest number of results. The method butterfly with $t_2 = {t_1}^{3/2}$ even improved the number of problems that SNOPT alone solved on average. Similarly, as in the previous experiment, the butterfly method benefits from the introduction of the parameter $s$ when using MINOS as a solver.

| **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 (1). The 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.

## 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 show that a specific relation between the parameters allows the method to converge to the desired M-stationary point. Additionally, in the particular case where MPCC-LICQ holds, 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 assuming any second-order conditions or strict complementarity-type conditions, which is an improvement over other methods.

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

Future research will focus on the main difficulty regarding relaxation schemes that are the convergence of approximate stationary sequences. A discussion regarding the above problem has been initiated in [7, 21] and appeal for further study.

**Acknowledgements** 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". The authors would like to thank the referees for their help and valuable comments.

# Appendix

## 7 Proof of a Technical Lemma

In the of proof Theorem 4 and Theorem 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 6** *Let $(I, I^c, I^-)$ be any partition of $\mathcal{I}^{00}_{GH}(x;t)$. Assume that the gradients*

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

*are linearly independent. Then, LICQ holds at $x$ for* (13)*.*

***Proof*** We show that the gradients of the constraints of (13) are positively linearly independent. For this purpose, we prove that the trivial solution is the only solution to the equation

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

where $\operatorname{supp}(\eta^g) \subseteq \mathcal{I}_g(x)$, $\operatorname{supp}(\eta^G) \subseteq \mathcal{I}_G(x;\bar{t})$, $\operatorname{supp}(\eta^H) \subseteq \mathcal{I}_H(x;\bar{t})$, $\operatorname{supp}(\eta^\Phi) \subseteq \mathcal{I}^{+0}_{GH}(x;t) \cup \mathcal{I}^{0+}_{GH}(x;t)$, $\operatorname{supp}(\nu^{F_1(x;t)}) \subseteq I$, $\operatorname{supp}(\nu^{F_2(x;t)}) \subseteq I$, $\operatorname{supp}(\mu^{F_1(x;t)}) \subseteq I^c$, $\operatorname{supp}(\mu^{F_2(x;t)}) \subseteq I^c$, $\operatorname{supp}(\delta^{F_1(x;t)}) \subseteq I^-$, $\operatorname{supp}(\delta^{F_2(x;t)}) \subseteq I^-$ where $I \cup I^c \cup I^- = \mathcal{I}^{00}_{GH}(x;t)$ 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.

We now replace those gradients in the equation above

$$0 = \sum_{i \in \mathcal{I}_g(x)} \lambda^g_i \nabla g_i(x) + \sum_{i=1}^{p} \lambda^h_i \nabla h_i(x) + \sum_{i=1}^{q} \lambda^G_i \nabla G_i(x) + \sum_{i=1}^{q} \lambda^H_i \nabla H_i(x),$$

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 $\theta$ and (8). Similarly, we get $\mu_i^{F_1(x;t)} = \mu_i^{F_2(x;t)} = \nu_i^{F_2(x;t)} = \nu_i^{F_1(x;t)} = 0$. □

## References


1. L. Abdallah, M. Haddou, and T. Migot. Solving absolute value equation using complementarity and smoothing functions. *Journal of Computational and Applied Mathematics*, 327:196–207, 2018.
2. L. Abdallah, M. Haddou, and T. Migot. A sub-additive dc approach to the complementarity problem. *Computational Optimization and Applications*, 73(2):509–534, 2019.
3. R. Andreani, G. Haeser, L. D. Secchin, and P. J.S. Silva. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM Journal on Optimization*, 29(4):3201–3230, 2019.
4. R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010.
5. M. S. Bazaraa and C. M. Shetty. *Foundations of optimization*, volume 122. Springer Science & Business Media, 2012.
6. V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes. A two-sided relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 16(2):587–609, 2005.

7. J.-P. Dussault, M. Haddou, A. Kadrani, and T. Migot. On approximate stationary points of the regularized mathematical program with complementarity constraints. *Journal of Optimization Theory and Applications*, 2020.
8. J.-P. Dussault, M. Haddou, and T. Migot. Mathematical programs with vanishing constraints: constraint qualifications, their applications, and a new regularization method. *Optimization*, 68(2-3):509–538, 2019.
9. M. L. Flegel and C. Kanzow. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005.
10. M. L. Flegel and C. Kanzow. On the guignard constraint qualification for mathematical programs with equilibrium constraints. *Optimization*, 54(6):517–534, 2005.
11. M. L. Flegel and C. Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints*. Springer, 2006.
12. R. Fletcher and S. Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004.
13. R. Fourer, D. Gay, and B. Kernighan. *Ampl*, volume 119. Boyd & Fraser, 1993.
14. P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005.
15. L. Guo, G.-H. Lin, and J. J. Ye. Solving mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 166(1):234–256, 2015.
16. T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1-2):257–288, 2013.
17. A. Kadrani, J.-P. Dussault, and A. Benchakroun. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 20(1):78–103, 2009.
18. C. Kanzow and A. 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. C. Kanzow and A. Schwartz. A New Regularization Method for Mathematical Programs with Complementarity Constraints with Strong Convergence Properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.
20. C. Kanzow and A. 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, 2014.
21. C. Kanzow and A. Schwartz. The Price of Inexactness: Convergence Properties of Relaxation Methods for Mathematical Programs with Complementarity Constraints Revisited. *Mathematics of Operations Research*, 40(2):253–275, 2015.
22. S. Leyffer. Macmpec: Ampl collection of mpecs. *Argonne National Laboratory. Available at www. mcs. anl. gov/leyfier/MacMPEC*, 2000.
23. G.-H. Lin and M. Fukushima. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, 133(1-4):63–84, 2005.
24. Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, 1996.
25. B. A. Murtagh and M. A. Saunders. Minos 5.0 user's guide. Technical report, DTIC Document, 1983.
26. J. V. Outrata. Optimality conditions for a class of mathematical programs with equilibrium constraints. *Mathematics of Operations Research*, 24(3):627–644, 1999.
27. J. V. Outrata. A generalized mathematical program with equilibrium constraints. *SIAM Journal on Control and Optimization*, 38(5):1623–1638, 2000.
28. J.-S. Pang and M. Fukushima. Complementarity constraint qualifications and simplified b-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications*, 13(1):111–136, 1999.

29. A. Ramos. Mathematical Programms with Equilibrium Constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization Online*, 2016.
30. H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.
31. S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.
32. A. Schwartz. *Mathematical programs with complementarity constraints: Theory, methods, and applications*. PhD thesis, Ph. D. dissertation, Institute of Applied Mathematics and Statistics, University of Würzburg, 2011.
33. S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.
34. A. Wächter and L. 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.
35. J. J. Ye. Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM Journal on Optimization*, 10(4):943–962, 2000.
36. J. J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005.
37. J. J. Ye and X. Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, 22(4):977–997, 1997.
38. J. J. Ye, D. L. Zhu, and Q. J. Zhu. Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM Journal on optimization*, 7(2):481–507, 1997.