

A Unified Framework of Regularization Methods for Degenerate Non-Linear Optimization Models

Tangi Migot, Jean-Pierre Dussault, Mounir Haddou

▶ **To cite this version:**

Tangi Migot, Jean-Pierre Dussault, Mounir Haddou. A Unified Framework of Regularization Methods for Degenerate Non-Linear Optimization Models. 2018. hal-01734832

HAL Id: hal-01734832
https://hal.science/hal-01734832

Preprint submitted on 15 Mar 2018

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.

A Unified Framework of Regularization Methods for Degenerate Non-Linear Optimization Models

Tangi Migot* Jean-Pierre Dussault† Mounir Haddou*

March 15, 2018

Abstract

In this paper, we study three classes of difficult non-linear optimization, problems with complementarity (MPCC), vanishing (MPVC) and cardinality constraints (OMCC). They all have in common degenerate constraints, which fail to satisfy classical constraint qualifications in a generic way. The feasible sets of these problems are non-convex, possibly non-connected, with an empty relative interior. This causes several difficulties in practice. In a recent work (Dussault et al. 2017), we propose a unified framework of methods that consider a regularization-penalization-active set method to solve the MPCC, which possesses the best known convergence properties. In this paper, we extend this unified framework to MPVC and OMCC and consider some applications on optimal control problems.

2010 MSC: 90C30, 90C33, 49M37, 65K05.
Keywords: Non-Linear Programming; MPCC; Cardinality Constraint; MPVC; Relaxation Methods.

1 Introduction

In this paper, we study a non-linear optimization model with degenerate constraints including complementarity constraints, vanishing constraints and cardinality constraints.

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & x \in \mathcal{X}, \end{aligned} \tag{1}$$

with

$$\mathcal{X} := \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} g(x) \leq 0, \\ 0 \leq G_1(x) \perp H_1(x) \geq 0, \\ H_2(x) \geq 0,\ G_2(x) \circ H_2(x) \leq 0, \\ H_3(x) \geq 0,\ G_3(x) \circ H_3(x) = 0, \end{array} \right\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $G_1, H_1 : \mathbb{R}^n \to \mathbb{R}^{q1}$, $G_2, H_2 : \mathbb{R}^n \to \mathbb{R}^{q2}$, $G_3, H_3 : \mathbb{R}^n \to \mathbb{R}^{q3}$. $\circ$ denotes the component-wise product of two vectors also known as Hadamard product.

We decide to skip classical equality constraints $h(x) = 0$ in order to simplify the presentation, although they could be successfully added without loss of generality in a straightforward way.

Problem (1) can be equivalently written as a special case of an optimization problem with geometric constraints

$$\min_{x\in\mathbb{R}^n} \ f(x) \text{ s.t. } F(x) \in \Gamma,$$

*Univ Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France. tangi.migot@insa-rennes.fr ; mounir.haddou@insa-rennes.fr

†Département d'Informatique, faculté des Sciences, Université de Sherbrooke, Canada. Jean-Pierre.Dussault@USherbrooke.CA ; This research was partially supported by NSERC grant.

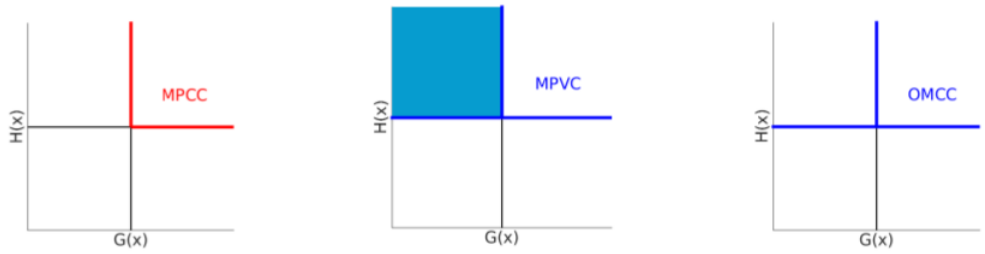


Figure 1: Feasible set of the complementarity, vanishing and kink constraint.

where $F(x) := (g(x), h(x), \Psi(x))$, $\Psi(x) := (G(x), H(x))$, $\Gamma :=]-\infty, 0]^p \times \{0\}^m \times C^q$) and $C := C_1 \cup C_2 \cup C_3 := \{(a,b) \ : \ 0 \leq a \perp b \geq 0\} \cup \{(a,b) \ : \ b \geq 0, \ ab \leq 0\} \cup \{(a,b) \ : \ b \geq 0, \ ab = 0\}$. In this context, we have $\mathcal{X} = F^{-1}(\Gamma)$.

This general form obviously includes Mathematical Programs with Complementarity Constraints (MPCC) [9, 26, 12, 14, 16, 17, 19, 20], Mathematical Programs with Vanshing Constraints (MPVC) [10, 15, 22, 1] and a more general form of Optimization Models with Cardinality Constraints (OMCC) [5, 6, 4, 7, 11] that we will call "Kink Constraints".

These 3 families of constraints are the most popular in the literature among the degenerate non-linear programs. MPVC and OMCC can both be cast as an MPCC. However, this approach leads to several difficulties as pointed out in [1] and in [6, 7].

Remark 1. *The motivation to consider kink constraints is to generalize what have been called cardinality constraints. In [6] the authors consider the relaxation of the cardinality constraint with*

$$\|x\|_0 \leq \kappa \Longleftrightarrow e^T y \geq n - \kappa, y \geq 0, x \circ y = 0.$$

In this case, the kink constraint is simplified, since the right-hand side of the degenerate constraint is an independent variable.

In this context solving the problem means finding a local minimum. Even so, this goal apparently modest is hard to achieve in general due to the degenerate nature of the MPCC. Therefore, numerical methods that consider only first order information may be expected to compute a stationary point.

The wide variety of approaches with this aim computes the KKT conditions, which requires some constraint qualifications to hold at the solution to be an optimality condition. However, it is well-known that these constraint qualifications never hold in general for (1). For instance, the classical Mangasarian-Fromowitz constraint qualification that is very often used to guarantee convergence of algorithms is violated at any feasible point. This is partly due to the geometry of the complementarity constraint that always has an empty relative interior.

These issues have motivated the definition of enhanced constraint qualifications and optimality conditions for the MPCC, MPVC and OMCC as in [17, 16, 29, 12] to cite some of the earliest research on MPCC. In particular, it was shown that the genuine necessary condition for these problems are M-stationary conditions.

In view of the constraint qualifications issues that plague the (1), 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 $g(x) \leq 0$ 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.

In [26], the authors introduce a unified framework of regularization methods for the MPCC, which contain most of the methods proposed in the literature with proved convergence to M-stationary points. Our motivation in this paper is to show the straightforward applicability of the unified framework for optimization methods (UFO) for the more general degenerate non-linear program (1).

2 Applications in Optimal Control Problems

In this section, we discuss some applications of degenerate non-linear programs applied to various optimal control problems. Complementarity constraints appear in a very natural way in many applications involving contact problem (for instance robot system with two modes: contact and no-contact. See optimal control of multiple robot systems with friction [27]) or change of phase, but also in many optimal control problems [3]. A large family of problems related with a very recent interest in the literature is the optimal control of sweeping process [8]. For instance, in [31] the authors study the quadratic optimal control problem with a linear complementarity system in the constraints.

2.1 Optimal control of the obstacle problem

One of the most typical example is the optimal control of the obstacle problem [24], which is a mathematical problem governed by variational inequalities in function space.

The distributed optimal control of the obstacle problem with control constraints [32]

$$\begin{aligned} \min_{y\in H_0^1(\Omega),u\in L^2(\Omega),\xi\in H^{-1}(\Omega)} \; & j(y)+\frac{\nu}{2}\|u\|^2_{L^2(\Omega)} \\ \text{s.t. } & \mathcal{A}y = u-\xi+f, \\ & 0\le \varphi - y \perp \xi \ge 0, \\ & u_a \le u \le u_b \text{ a.e. in } \Omega. \end{aligned} \tag{2}$$

Typical assumptions suppose $\Omega \subset \mathbb{R}^n$ $(n \geq 1)$ open and bounded, $j(y)$ a Fréchet-differentiable observation term $j : H_0^1(\Omega) \to \mathbb{R}$ of the state y of an $L^2(\Omega)$ regularization term with $\nu > 0$. The bounded linear operator $\mathcal{A} : H_0^1(\Omega) \to H^{-1}(\Omega)$ is assumed to be coercive. The right-hand side f belongs to $H^{-1}(\Omega)$. The control bounds satisfy $u_a, u_b \in H^1(\Omega)$. The obstacle $\varphi \in H^1(\Omega)$ satisfies $\varphi \geq 0$ on Γ in the sense that $\min(\varphi, 0) \in H_0^1(\Omega)$.

Existence of minimizers of (2) can be shown assuming j to be bounded below and weakly lower semicontinuous [25, Thm 2.1]. The study of stationary conditions of (2) has been focused on S-stationary conditions assuming additional assumptions on the set of admissible controls. In [24], the authors consider a discretized version of this problem to prove existence of solutions and study the stationary properties of this problem. The satisfaction of this condition had some practical interest for instance in [24] to prove convergence of the discretization scheme and discretization error estimates.

2.2 Locomotion problem

The locomotion problem is to move from a given start to a given end position without considering individual steps as a human would do them. This macroscopic perspective considers a plant with continuous dynamics that can be described by ordinary differential equations. If combined with a suitable cost function one obtains a standard optimal control problem. The direct approach to optimal control is chosen here and thus a combination of a discretization technique and a non-linear optimization method is used. The goal of the considered inverse optimal control task is to determine a cost function within a given parametrized family of cost functions such that the corresponding optimal control result has minimal distance to given data. In consequence, this problem is a special bilevel optimal control problem where the lower level is the optimal control problem and the upper level is the inversion problem.

In [2], the authors consider numerical methods to solve this problem using relaxation method SS and SU as well as some lifting approach to solve the corresponding MPCC.

2.3 Robot motion planning

The complementarity constraint is the most popular degenerate constraint, but the vanishing constraint also appears in a very natural way, since it can be used to express logic implication, i.e. for some index i

$$(G_i(x)H_i(x) \leq 0, H_i(x) \geq 0) \Longleftrightarrow (0 < H_i(x) \Longrightarrow G_i(x) \leq 0)\,.$$

One example of logic constraints in a real-world application arises in robot motion planning [23]. Here, a communication network of a given density needs to be maintained among a swarm of independent mobile robots. For each pair (i, j) of robots, $H_{i,j}(x) > 0$ indicates that the pair is communicating. Then, $0 \leq G_{i,j}(x)$ must be satisfied to ensure that the distance between robots i and j actually allows for communication. Conversely, this distance constraint vanishes for each pair (i, j) of robots with $H_{i,j}(x) = 0$ which do not communicate. The study of an active-set approach for the related MPVC as well as numerical results have been shown in [22].

3 Preliminaries

In this section, we introduce some definitions, some preliminary notions and their consequences.

Given $x^* \in \mathcal{X}$, we denote

$$\begin{aligned}
\mathcal{I}^{+0} &:= \{i \mid G_i(x^*) > 0 \text{ and } H_i(x^*) = 0\},\\
\mathcal{I}^{-0} &:= \{i \mid G_i(x^*) < 0 \text{ and } H_i(x^*) = 0\},\\
\mathcal{I}^{\pm 0} &:= \mathcal{I}^{+0} \cup \mathcal{I}^{-0},\\
\mathcal{I}^{-+} &:= \{i \mid G_i(x^*) < 0 \text{ and } H_i(x^*) > 0\},\\
\mathcal{I}^{0+} &:= \{i \mid G_i(x^*) = 0 \text{ and } H_i(x^*) > 0\},\\
\mathcal{I}^{00} &:= \{i \mid G_i(x^*) = 0 \text{ and } H_i(x^*) = 0\}.
\end{aligned}$$

Furthermore, we denote

$$\begin{aligned}
\mathcal{I}_{CC} &:= \{1, \ldots, q1\},\\
\mathcal{I}_{VC} &:= \{q1 + 1, \ldots, q1 + q2\},\\
\mathcal{I}_{KC} &:= \{q1 + q2 + 1, \ldots, q1 + q2 + q3\},
\end{aligned}$$

and

$$q = q1 + q2 + q3.$$

In a comprehensive way, we denote $\mathcal{I}^{00}_{CC} = \mathcal{I}^{00} \cap \mathcal{I}_{CC}$.

The polar cone of a cone K is a closed and convex cone defined by $K^\circ := \{d \mid d^T x \leq 0\, \forall x \in K\}$.

The tangent cone of a set Ω at $x^* \in \Omega$ is a closed cone defined by

$$\mathcal{T}_\Omega(x^*) := \{d \mid d = \lim_{k\to\infty} t_k(x^k - x^*) \text{ with } t_k \geq 0 \text{ and } x^k \to x^* \text{ with } x^k \in \Omega\}.$$

The regular (or Fréchet) normal cone of a set Ω at $x^* \in \Omega$ is a closed cone defined by

$$\hat{\mathcal{N}}_\Omega(x^*) := \{d \mid d^T(x - x^*) \leq o(\|x - x^*\|)\ \forall x \in \Omega\}.$$

The limiting (or Mordukhovich) normal cone of a set Ω at $x^* \in \Omega$ is a closed cone defined by

$$\mathcal{N}_\Omega(x^*) := \{d \mid d = \lim_{k\to\infty} d^k \text{ with } d^k \in \hat{\mathcal{N}}_\Omega(x^k) \text{ and } x^k \to x^* \text{ with } x^k \in \Omega\}.$$

We consider the specialized sets

$$\begin{aligned}
\mathcal{L}_{MPCC}(x^*) := \{d \mid 0 \leq \nabla G_{1i}(x^*)^T d \perp \nabla H_{1i}(x^*)^T d \geq 0\ (i \in \mathcal{I}^{00}),\\
\nabla G_{1i}(x^*)^T d \geq 0\ (i \in \mathcal{I}^{+0}),\ \nabla H_{1i}(x^*)^T d \geq 0\ (i \in \mathcal{I}^{0+})\},
\end{aligned}$$

$$\mathcal{L}_{MPVC}(x^*) := \{d \mid \nabla H_{2i}(x^*)^T d = 0 \ (i \in \mathcal{I}^{+0}), \nabla H_{2i}(x^*)^T d \geq 0 \ (i \in \mathcal{I}^{00} \cup \mathcal{I}^{-0}),$$
$$\nabla G_{2i}(x^*)^T d \leq 0 \ (i \in \mathcal{I}^{0+}), (\nabla H_{2i}(x^*)^T d)(\nabla G_{2i}(x^*)^T d) \leq 0 \ (i \in \mathcal{I}^{00})\},$$

and

$$\mathcal{L}_{OMCC}(x^*) := \{d \mid d^T \nabla H_{3i}(x^*) = 0 \ (i \in \mathcal{I}^{\pm 0}), d^T \nabla H_{3i}(x^*) \geq 0 \ (i \in \mathcal{I}^{00}),$$
$$d^T \nabla G_{3i}(x^*) = 0 \ (i \in \mathcal{I}^{0+}), (d^T \nabla G_{3i}(x^*))(d^T \nabla H_{3i}(x^*)) = 0 \ (i \in \mathcal{I}^{00})\}.$$

The generalized linearized cone of (1) at $x^* \in \mathcal{X}$ is the closed cone defined by

$$\mathcal{L}_{\mathcal{X}} := \{d \mid \nabla g(x^*)^T d \leq 0\} \cap \mathcal{L}_{MPCC}(x^*) \cap \mathcal{L}_{MPVC}(x^*) \cap \mathcal{L}_{OMCC}(x^*).$$

Note that this cone is not polyhedral and is in general not convex. However, we denote it linearized cone since it can be interpreted as a first order approximation of the tangent cone of $\mathcal{X}$ at x^*.

Definition 3.1. *A point $x^* \in \mathcal{X}$ is said B-stationary if*

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

We see from this definition that a B-stationary point is closely linked with local minima of a linearization of (1). This is formalized in the following result.

Theorem 3.2. *Let $x^* \in \mathcal{X}$ be a local minimum of (1) that satisfies the following constraint qualification:*

$$\mathcal{T}_{\mathcal{X}}(x^*)^\circ = \mathcal{L}_{\mathcal{X}}(x^*)^\circ. \tag{3}$$

Then, x^* is a B-stationary point.

Proof. It follows from local optimality of x^* that

$$\nabla f(x^*)^T d \geq 0, \ \forall d \in \mathcal{T}_{\mathcal{X}}(x^*),$$

which implies

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

Using the constraint qualification (3), x^* is then a B-stationary point. □

Obviously, computing a B-stationary point is already a very hard problem. Thus, in the next section, we define some other optimality conditions that are more tractable from a computational point of view.

4 Optimality Conditions and Constraint Qualifications

We introduce two conditions that we will show are interesting optimality conditions for this problem. We use here the following notation $\nabla F(x^*)\mathcal{N}_{\mathcal{X}}(F(x^*)) = \{\nabla F(x^*)\nu \mid \nu \in \mathcal{N}_{\mathcal{X}}(F(x^*))\}$.

Definition 4.1. *A point $x^* \in \mathcal{X}$ is said*

- *M-stationary if $-\nabla f(x^*) \in \nabla F(x^*)\mathcal{N}_{\mathcal{X}}(F(x^*))$;*
- *S-stationary if $-\nabla f(x^*) \in \nabla F(x^*)\hat{\mathcal{N}}_{\mathcal{X}}(F(x^*))$.*

We observe that S-stationarity implies M-stationarity.

In the special case where $q1 + q2 + q3 = 0$, the problem (1) is reduced to a classical non-linear program. In this case, it is well-known that under a classical constraint qualification, for instance Guignard Constraint Qualification (GCQ) $\mathcal{T}_{\mathcal{X}}(x^*)^\circ = \mathcal{L}_{\mathcal{X}}(x^*)^\circ$, a good optimality condition is given by the KKT condition. We say that $x^* \in \mathcal{X}$ is a KKT point if there exists $\lambda^g \in \mathbb{R}^p_+$ with $\lambda^G_i = 0 \ \forall i \notin \mathcal{I}_g(x^*)$ such that $-\nabla f(x^*) = \nabla g(x^*)\lambda^g$.

The following result shows the link between S-stationary points and KKT points of (1). The Lagrangian function associated to (1) at x^* is given by

$$L(x^*) := f(x^*) + \lambda^g g(x^*) - \nu^G \nabla G(x^*) - \nu^H \nabla H(x^*),$$

where $(\lambda, \nu^G, \nu^H) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$.

Theorem 4.2.

$$x^* \text{ is a KKT point of } (1) \iff x^* \text{ is an S-stationary point.}$$

Furthermore, this condition can be explicitly written as

$$\begin{cases} \nabla L(x^*) = 0, \\ \lambda_i^g = 0 \ (i \notin \mathcal{I}_g), \lambda_i^g \geq 0 \ (i \in \mathcal{I}_g), \\ \nu_i^G = 0 \ (i \in \mathcal{I}_{CC}^{+0}), \ \nu_i^H = 0 \ (i \in \mathcal{I}_{CC}^{0+}), \\ \nu_i^G \geq 0, \ \nu_i^H \geq 0 \ (i \in \mathcal{I}_{CC}^{00}). \\ \nu_i^G = 0 \ (i \in \mathcal{I}_{VC}^{\pm 0} \cup \mathcal{I}_{VC}^{-+}), \ \nu_i^G \leq 0 \ (i \in \mathcal{I}_{VC}^{0+}), \\ \nu_i^H = 0 \ (i \in \mathcal{I}_{VC}^{0+} \cup \mathcal{I}_{VC}^{-+}), \ \nu_i^H \geq 0 \ (i \in \mathcal{I}_{VC}^{-0}) \\ \nu_i^H \geq 0, \nu_i^G = 0 \ (i \in \mathcal{I}_{VC}^{00}), \\ \nu_i^H = 0 \ (i \in \mathcal{I}_{KC}^{0+}), \ \nu_i^G = 0 \ (i \in \mathcal{I}_{KC}^{\pm 0}), \\ \nu_i^G = 0, \nu_i^H \geq 0 \ (i \in \mathcal{I}_{KC}^{00}). \end{cases} \tag{4}$$

Proof. The condition (4) is obtained in a straightforward way from the definition of KKT point. It remains to show that it is equivalent to the Definition 4.1.

We know that for a given finite collection of sets $\Gamma = \cap_i \Gamma_i \subset \mathbb{R}^n$ it holds true that $\hat{\mathcal{N}}_\Gamma(x^*) = \cap_i \hat{\mathcal{N}}_{\Gamma_i}(x^*)$ by [28, Proposition 6.41]. Thus, it is sufficient to compute $\hat{\mathcal{N}}_{C_1}(x^*)$,$\hat{\mathcal{N}}_{C_2}(x^*)$ and $\hat{\mathcal{N}}_{C_3}(x^*)$:

- For C_1, it has been pointed out in [14, Proposition 2.1] that

$$\hat{\mathcal{N}}_{C_1}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{c} d_1 = 0, d_2 \in \mathbb{R} \text{ if } a = 0 < b, \\ d_1 \in \mathbb{R}, d_2 = 0 \text{ if } a > 0 = b, \\ d_1 \leq 0, d_2 \leq 0 \text{ if } a = b = 0 \end{array} \right\},$$

- For C_2, direct computation gives

$$\hat{\mathcal{N}}_{C_2}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{c} d_1 = 0, d_2 \leq 0 \text{ if } a < 0 = b, \\ d_1 = 0, d_2 \in \mathbb{R} \text{ if } a > 0 = b, \\ d_1 \geq 0, d_2 = 0 \text{ if } a = 0 < b, \\ d_1 = 0, d_2 = 0 \text{ if } a < 0 < b, \\ d_1 = 0, d_2 \leq 0 \text{ if } a = b = 0 \end{array} \right\},$$

- For C_3, direct computation gives

$$\hat{\mathcal{N}}_{C_3}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{c} d_1 \in \mathbb{R}, d_2 = 0 \text{ if } a = 0 < b, \\ d_1 = 0, d_2 \in \mathbb{R} \text{ if } a \neq 0 = b, \\ d_1 = 0, d_2 \leq 0 \text{ if } a = b = 0 \end{array} \right\}.$$

This concludes the proof.

□

In the special case of $q1 = q2 = 0$, i.e. only OMCC, with $n_1 = n - q3$ such that $x, y \in \mathbb{R}^{n_1} \times \mathbb{R}^{q3}$ and $G_3(x,y) = x$, $H(x,y) = y$ and $f(x,y) = f(x)$ we get:

$$\begin{aligned} &-\nabla f(x^*) = \lambda^g \nabla g(x^*) - \nu^G \nabla G(x^*) - \nu^H \nabla H(x^*), \\ &\lambda_i^g = 0 \ (i \notin \mathcal{I}_g(x^*)), \ \nu^H = 0 \ (i \in \mathcal{I}^{0+}), \ \nu^H \geq 0 \ (i \in \mathcal{I}^{00}), \ \nu^G = 0 \ (i \in \mathcal{I}^{00} \cup \mathcal{I}^{+0} \cup \mathcal{I}^{-0}). \end{aligned} \tag{5}$$

This condition corresponds to the definition of S-stationary point for OMCC in [6]. In the mentioned paper, the authors prove that under a weak constraint qualification a local minimum of the problem satisfies (5). However, this is no longer true in our more general context of kink constraints as illustrated by the following example.

Example 1. *Consider the following three dimensional example:*

$$\begin{aligned}&\min_{x\in\mathbb{R}^3} x_2 - x_3\\ &s.t.\ x_3 - 4x_1 \le 0, x_3 - 4x_2 \le 0,\\ &\quad x_2 \ge 0, x_1 \circ x_2 = 0.\end{aligned}$$

S-stationary condition for this problem at the local minimum $x^* = (0,0,0)^T$ *yields to*

$$\begin{aligned}&\begin{pmatrix} 0\\ -1\\ 1\end{pmatrix} = \lambda_1 \begin{pmatrix} -4\\ 0\\ 1\end{pmatrix} + \lambda_2 \begin{pmatrix} 0\\ -4\\ 1\end{pmatrix} - \nu^G \begin{pmatrix} 1\\ 0\\ 0\end{pmatrix} - \nu^H \begin{pmatrix} 0\\ 1\\ 0\end{pmatrix},\\ &\lambda_1 \ge 0, \lambda_2 \ge 0, \nu^G = 0, \nu^H \ge 0.\end{aligned}$$

However, the only solution in the above equation is

$$\lambda_1 = 0, \lambda_2 = 1, \nu^G = 0, \nu^H = -3,$$

which is a contradiction with the sign of the multiplier ν^H*.*

Similar examples can be found for MPCC and MPVC in the literature. This phenomenon is generic from degenerate non-linear program as (1).

Theorem 4.3. *The M-stationarity condition at* $x^* \in \mathcal{X}$ *is equivalent to*

$$\begin{cases} \nabla L(x^*) = 0,\\ \lambda_i^g = 0\ (i \notin \mathcal{I}_g), \lambda_i^g \ge 0\ (i \in \mathcal{I}_g),\\ \nu_i^G = 0\ (i \in \mathcal{I}_{CC}^{+0}),\ \nu_i^H = 0\ (i \in \mathcal{I}_{CC}^{0+}),\\ \nu_i^G \nu_i^H = 0\ or\ \nu_i^G > 0, \nu_i^H > 0\ (i \in \mathcal{I}_{CC}^{00}),\\ \nu_i^G = 0\ (i \in \mathcal{I}_{VC}^{\pm 0} \cup \mathcal{I}_{VC}^{-+}),\ \nu_i^G \le 0\ (i \in \mathcal{I}_{VC}^{0+}),\\ \nu_i^H = 0\ (i \in \mathcal{I}_{VC}^{0+} \cup \mathcal{I}_{VC}^{-+}),\ \nu_i^H \ge 0\ (i \in \mathcal{I}_{VC}^{-0})\\ \nu_i^G \nu_i^H = 0, \nu_i^G \le 0\ (i \in \mathcal{I}_{VC}^{00}),\\ \nu_i^H = 0\ (i \in \mathcal{I}_{KC}^{0+}),\ \nu_i^G = 0\ (i \in \mathcal{I}_{KC}^{\pm 0}),\\ \nu_i^G \nu_i^H = 0\ (i \in \mathcal{I}_{KC}^{00}).\end{cases} \tag{6}$$

Proof. We proceed in a similar way as in Theorem 4.2.

We know that for a given finite collection of sets $\Gamma = \cap_i \Gamma_i \subset \mathbb{R}^n$ it holds true that $\mathcal{N}_\Gamma(x^*) = \cap_i \mathcal{N}_{\Gamma_i}(x^*)$ by [28, Proposition 6.41]. Thus, it is sufficient to compute $\mathcal{N}_{C_1}(x^*)$,$\mathcal{N}_{C_2}(x^*)$ and $\mathcal{N}_{C_3}(x^*)$. It can be noted here that these sets only differ for indices $i \in \mathcal{I}^{00}$ compared to those from the proof of Theorem 4.2.

- For C_1, it has been proved in [14, Proposition 2.1] that

$$\mathcal{N}_{C_1}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{c} d_1 = 0, d_2 \in \mathbb{R} \text{ if } a = 0 < b,\\ d_1 \in \mathbb{R}, d_2 = 0 \text{ if } a > 0 = b,\\ \text{either } d_1 d_2 = 0 \text{ or } d_1 \le 0, d_2 \le 0 \text{ if } a = b = 0 \end{array} \right\}.$$

- For C_2, direct computation gives

$$\hat{\mathcal{N}}_{C_2}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{c} d_1 = 0, d_2 \le 0 \text{ if } a < 0 = b,\\ d_1 = 0, d_2 \in \mathbb{R} \text{ if } a > 0 = b,\\ d_1 \ge 0, d_2 = 0 \text{ if } a = 0 < b,\\ d_1 = 0, d_2 = 0 \text{ if } a < 0 < b,\\ d_1 d_2 = 0, d_1 \ge 0 \text{ if } a = b = 0 \end{array} \right\},$$

- For C_3, direct computation gives

$$\hat{\mathcal{N}}_{C_3}(a,b) = \left\{ (d_1,d_2) \;\middle|\; \begin{array}{l} d_1 \in \mathbb{R}, d_2 = 0 \text{ if } a = 0 < b, \\ d_1 = 0, d_2 \in \mathbb{R} \text{ if } a \neq 0 = b, \\ \quad d_1 d_2 = 0 \text{ if } a = b = 0 \end{array} \right\}.$$

This concludes the proof.

□

The observation that S-stationarity may not hold at a local minimum has motivated the following key result.

Theorem 4.4. *Let $x^* \in \mathcal{X}$ be a local minimum of* (1) *that satisfies the following constraint qualification:*

$$\mathcal{T}_{\mathcal{X}}(x^*)^\circ \subset \nabla F(x^*)\mathcal{N}_{\mathcal{X}}(F(x^*)). \tag{7}$$

Then, x^* is an M-stationary point. Conversely, if $x^* \in \mathcal{X}$ is an M-stationary point, then (7) *holds.*

Proof. The first part of the proof follows the exact same steps of Theorem 3.2.

We now prove the converse. Let x^* be an M-stationary point. For any $d \in \mathcal{T}_{\mathcal{X}}(x^*)$, by [13], there exists a smooth function φ such that x^* is a local optimum of $\min_{x\in\mathcal{X}} \varphi(x)$ and $-\nabla\varphi(x^*) = d$. It follows that x^* is an M-stationary point of $\min_{x\in\mathcal{X}} \varphi(x)$ and hence, by Definition 4.1, we have

$$d = -\nabla\varphi(x^*) \subset \nabla F(x^*)\mathcal{N}_{\mathcal{X}}(F(x^*)).$$

Since, we choose arbitrarily d, we have $\mathcal{T}_{\mathcal{X}}(x^*)^\circ \subset \nabla F(x^*)\mathcal{N}_{\mathcal{X}}(F(x^*))$. □

We know from [14] and [10] respectively for MPCC and MPVC that the condition (7) is the weakest constraint qualification that ensures such result. Besides, in [10] the authors give an explicit formulation of this condition. It is to be noted that both constraint qualifications in Theorem 3.2 and Theorem 4.4 are independent from the choice of the objective function.

We conclude this section by showing that condition (3) given in Theorem 3.2 is stronger than condition (7) given in Theorem 4.4.

Theorem 4.5. *Let $x^* \in \mathcal{X}$ satisfies* (3)*, then it also satisfies* (7)*.*

Example 3.1 in [14] in the context of MPCC shows that this result is sharp, since we do not have the equality in general.

Remark 2. *Another degenerate kind of constraint that could have been considered is the "cross constraint", i.e.*

$$\{x \mid G(x) \circ H(x) = 0\}.$$

In this case, a remarkable phenomenon arises, since the Fréchet normal cone is then reduced to the singleton 0, while the limiting normal cone gives $\{(d_1,d_2) \mid d_1 d_2 = 0\}$. So, the S-stationary condition does not give any pertinent information.

5 A Constraint Qualification to Show Convergence of the Regularization Methods

In the previous sections, we introduced very weak constraint qualifications that have been used to provide optimality conditions. In particular, Theorem 4.4 shows that our goal is to define a numerical method, which converges to M-stationary points. However, both constraint qualifications are very hard to check in practice, and they may not be sufficient to prove useful algorithmic properties. In this section, we introduce a new constraint qualification and prove that it is useful in an algorithmic context.

Definition 5.1. *G-GCRSC holds at x^* if for any partition*

$$\begin{aligned}\mathcal{I}^{00}_{CC} &= A_{CC} \cup B_{CC} \cup C_{CC},\\ \mathcal{I}^{00}_{VC} &= A_{VC} \cup B_{VC},\\ \mathcal{I}^{00}_{KC} &= A_{KC} \cup B_{KC},\end{aligned}$$

such that

$$\sum_{i=1}^{p} \nu_i^g \nabla g_i(x^*) - \sum_{i=1}^{q} \nu_i^G \nabla G_i(x^*) - \sum_{i=1}^{q} \nu_i^H \nabla H_i(x^*) = 0,$$

with

$$\begin{aligned}&\nu_i^g = 0 \ (i \notin \mathcal{I}_g),\\ &\nu_i^G = 0 \ (i \notin \mathcal{I}^{0+}_{CC} \cup A_{CC} \cup \mathcal{I}^{0+}_{VC} \cup A_{VC} \cup \mathcal{I}^{0+}_{KC} \cup A_{KC}),\\ &\nu_i^H = 0 \ (i \notin \mathcal{I}^{+0}_{CC} \cup A_{CC} \cup \mathcal{I}^{\pm 0}_{VC} \cup B_{VC} \cup \mathcal{I}^{\pm 0}_{KC} \cup B_{KC}),\end{aligned}$$

and $\nu_i^g \geq 0$ $(i \in \mathcal{I}_g(x^*))$, ν_i^G and $\nu_i^H \geq 0$ $(i \in A_{CC})$, $\nu_i^G > 0$ $(i \in C_{CC})$, ν_i^H $(i \in B_{CC}) > 0$, $\nu_i^G \leq 0$ $(i \in \mathcal{I}^{0+}_{VC} \cup A_{VC})$, $\nu_i^H \geq 0$ $(i \in \mathcal{I}^{-0}_{VC})$, there exists $\delta > 0$ such that the family of gradients

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

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

$$\begin{aligned}&\mathcal{I}_1 := \{i \in \mathcal{I}_g(x^*) | -\nabla g_i(x^*) \in \mathcal{P}\},\\ &\mathcal{I}_3 := \mathcal{I}^{0+}_{CC} \cup C_{CC} \cup \mathcal{I}^{0+}_{KC} \cup A_{KC} \cup \{i \in A_{CC} \cup \mathcal{I}^{0+}_{VC} \cup A_{VC} | \nabla G_i(x^*) \in \mathcal{P}\},\\ &\mathcal{I}_4 := \mathcal{I}^{+0} \cup B \cup \mathcal{I}^{-0}_{KC} \cup \{i \in A_{CC} | \nabla H_i(x^*) \in \mathcal{P}\} \cup \{i \in \mathcal{I}^{-0}_{VC} | -\nabla H_i(x^*) \in \mathcal{P}\},\end{aligned}$$

with the notations

$$\mathcal{P} = \nabla F(x^*) \mathcal{N}_{\mathcal{X}}(F(x^*)) \text{ and } B = B_{CC} \cup B_{VC} \cup B_{KC}.$$

In the special case where there is no partition of $\mathcal{I}^{00}$ that satisfies the condition of the definition above, the gradients are obviously linearly independent (G-LICQ).

Furthermore, G-GCRSC is weaker than assuming constant rank of the family of gradients of active constraints in a neighborhood (G-CRCQ), since the G-GCRSC condition considers only the family of gradients that are linearly dependent with coefficients that have M-stationary signs.

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 5.1. *Let $\{x^k\}$ and $0 \neq \{\nu^k\} \in \mathbb{R}^p_+ \times \mathbb{R}^{q1+q2+q3} \times \mathbb{R}^{q1+q2+q3}$ be such that $x^k \to x^*$ and*

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

with $\nu^{g,k} \geq 0$ $(i \in \mathcal{I}_g)$, $\nu_i^G \leq 0$ $(i \in \mathcal{I}^{0+}_{VC})$,$\nu_i^H \geq 0$ $(i \in \mathcal{I}^{-0}_{VC})$.

(ii) $$\begin{cases} \lim\limits_{k\to\infty} \frac{\nu_i^{g,k}}{\|\nu^k\|_\infty} = 0 \ (i \notin \mathcal{I}_g), \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}}{\|\nu^k\|_\infty} = 0 \ (i \notin \mathcal{I}^{0+}_{CC} \cup \mathcal{I}^{0+}_{VC} \cup \mathcal{I}^{0+}_{KC}) \\ \text{and } \lim\limits_{k\to\infty} \frac{\nu_i^{H,k}}{\|\nu^k\|_\infty} = 0 \ (i \notin \mathcal{I}^{+0}_{CC} \cup \mathcal{I}^{\pm 0}_{VC} \cup \mathcal{I}^{\pm 0}_{KC}), \end{cases}$$

(iii)
$$\begin{cases} \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}\nu_i^{H,k}}{\|\nu^k\|_\infty^2} = 0 \ or \ \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}}{\|\nu^k\|_\infty} > 0, \lim\limits_{k\to\infty} \frac{\nu_i^{H,k}}{\|\nu^k\|_\infty} > 0 \ (i \in \mathcal{I}_{CC}^{00}), \\ \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}\nu_i^{H,k}}{\|\nu^k\|_\infty^2} = 0 \ or \ \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}}{\|\nu^k\|_\infty} < 0 \ (i \in \mathcal{I}_{VC}^{00}), \\ \lim\limits_{k\to\infty} \frac{\nu_i^{G,k}\nu_i^{H,k}}{\|\nu^k\|_\infty^2} = 0 \ (i \notin \mathcal{I}_{KC}^{00}), \end{cases}$$

(iv) *the family of gradients of non-vanishing multipliers in* (i) *are linearly independent.*

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

Remark 3. *Note that assumption* (iv) *is not restrictive. According to [30, Lemma 7.1], we can build a sequence of multipliers that satisfies* (i), (ii) *and* (iii), *such that the gradients corresponding to non-vanishing multipliers in equation* (i) *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.*

The first step in our analysis is to prove that the sequences of multipliers satisfying Assumption 5.1 are bounded.

Theorem 5.2. *Given two sequences* $\{x^k\}$,$\{\nu^k\}$ *that satisfy Assumption 5.1. Suppose that* $x^k \to x^* \in \mathcal{X}$, *and G-CRSC holds at* x^*. *Then, the sequence* $\{\nu^k\}$ *is bounded.*

The proof is skipped here, since it is a straightforward extension of the ones presented for MPCC and MPVC respectively in [9] and [10].

A major consequence of the previous result is now stated.

Corollary 1. *Given two sequences* $\{x^k\}$,$\{\nu^k\}$ *that satisfy Assumption 5.1. Suppose that* $x^k \to x^* \in \mathcal{X}$, *and G-CRSC holds at* x^*. *Then,* x^* *is an M-stationary point of* (1).

Up to this point, it could be noticed that boundedness is not necessary to obtain Corollary 1, however, it will be of importance in the study of the convergence of the relaxation method.

6 UFO : Unified Framework of Optimization Methods

6.1 Regularization Methods for (1)

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

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

with $\bar{t}^G, \bar{t}^H : \mathbb{R}_+^l \to (\mathbb{R}_+ \cup \{+\infty\})^{q1+q2+q3}$ such that $\bar{t}_i^G = \infty$ for $i \in \mathcal{I}_{VC} \cup \mathcal{I}_{KC}$ and $\bar{t}(t) = \bar{t}^G_{\mathcal{I}_{CC}} = \bar{t}^H_{\mathcal{I}_{CC}\cup\mathcal{I}_{VC}\cup\mathcal{I}_{KC}}$ and $\bar{t} : \mathbb{R}_+^l \to \mathbb{R}_+$ such that $\lim_{\|t\|\to 0} \bar{t}(t) \to 0$ and the relaxation map $\Phi : \mathbb{R}^n \to \mathbb{R}^{q1+q2+2q3}$.

In the sequel we skip the dependency in t and denote $\bar{t}$ to simplify the notation. It is to be noted here that t is a vector of an arbitrary size denoted l as for instance in [9] where $l = 2$.

Remark 4. *Note here that the length of the relaxation map has been augmented due to the cardinality constraints, which requires a special care. We make the following assumption for* $i \in \mathcal{I}_{KC}$ *that the cardinality constraint is relaxed through two relaxation maps:*

$$\Phi_i(G(x), H(x); t) \le 0, \ \Phi_i(-G(x), H(x); t) \le 0. \tag{8}$$

The generalized Lagrangian function of $(R_t(x))$ is defined for $\eta \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{q1+q2+q3} \times \mathbb{R}^{q1+q2+q3} \times \mathbb{R}^{q1+q2+2q3}$ as

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

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

$$\mathcal{I}_\Phi(x; t) := \{i \mid \Phi_i(G(x), H(x); t) = 0\} \cup \{i \mid \Phi_i(-G(x), H(x); t) = 0\}.$$

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

- $\Phi(G(x), H(x); t)$ is a continuously differentiable real valued map extended component by component, so that
$$\Phi_i(G(x), H(x); t) := \Phi(G_i(x), H_i(x); t). \tag{H1}$$
- Direct computations give that the gradient with respect to x for $i \in \{1, \ldots, q1 + q2 + 2q3\}$ of $\Phi_i(G(x), H(x); t)$ for all $x \in \mathbb{R}^n$ is given by
$$\nabla_x\Phi_i(G(x), H(x); t) = \nabla G_i(x)\alpha^H_i(x; t) + \nabla H_i(x)\alpha^G_i(x; t),$$
where $\alpha^H(x; t)$ and $\alpha^G(x; t)$ are continuous maps by smoothness assumption on $\Phi(G(x), H(x); t)$, which we assume satisfy $\forall x \in \mathcal{X}$
$$\lim_{\|t\|\to 0} \alpha^H(x; t) = H(x) \text{ and } \lim_{\|t\|\to 0} \alpha^G(x; t) = G(x). \tag{H2}$$
- At the limit when $\|t\|$ goes to 0, the feasible set of the non-linear parametric program $(R_t(x))$ must converge to the feasible set of (1). In other words, given $\mathcal{F}(t)$ the feasible set of $(R_t(x))$ it holds that
$$\lim_{\|t\|\to 0} \{x \mid \Phi(G(x), H(x); t) \leq 0\} = \{x \mid G(x) \circ H(x) \leq 0\}, \tag{H3}$$
where the limit is assumed pointwise.
- At the boundary of the feasible set of the relaxation of the complementarity constraint it holds that for all $i \in \{1, \ldots, q1 + q2 + 2q3\}$
$$\Phi_i(G(x), H(x); t) = 0 \Longrightarrow F_{Gi}(x; t) = 0 \text{ or } F_{Hi}(x; t) = 0, \tag{H4}$$
where
$$\begin{aligned} F_G(x; t) &:= G(x) - \psi(H(x); t), \\ F_H(x; t) &:= H(x) - \psi(G(x); t), \end{aligned} \tag{9}$$
and ψ is a continuously differentiable real valued function extended component by component. Note that the function ψ may be two different functions in (9) as long as they satisfy the assumptions below. Those functions $\psi(H(x); t)$, $\psi(G(x); t)$ are non-negative for all $x \in \{x \in \mathbb{R}^n \mid \Phi(G(x), H(x); t) = 0\}$ and satisfy $\forall z \in \mathbb{R}^{q1+q2+q3}$
$$\lim_{\|t\|\to 0} \psi(z; t) = 0. \tag{H5}$$

Remark 5. *These assumptions are not restrictive and motivated by the fact that $\Phi(G(x), H(x); t)$ is an approximation (at the limit) of $G(x) \circ H(x)$ as shown by* (H3). (H1) *is coherent with the C^1 assumptions on all the functions involved in* (1). *By construction of Φ and regularity assumption,* (H2) *is logical. Finally,* (H4) *gives a structure of the relaxation map, where* (H5) *is coherent with* (H3).

Remark 6. *According to* (8)*, we also consider* $\Phi_i(-G(x), H(x); t)$*. We clarify here how the hypotheses extend to this case. The gradient of the relaxation map in* (H2) *becomes*

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

while in (H4) *we get*

$$\Phi_i(G(x), H(x); t) = 0 \Longrightarrow -F_{Gi}(x; t) = 0 \text{ or } F_{Hi}(x; t) = 0.$$

We will prove in Theorem 6.4 that this generic relaxation scheme for (1) converges to an M-stationary point requiring the following essential assumption on the functions ψ. As t goes to 0 the derivative with respect to the first variable of ψ satisfies $\forall z \in \mathbb{R}^{q1+q2+q3}$

$$\lim_{\|t\|\to 0} \left.\frac{\partial \psi(x; t)}{\partial x}\right|_{x=z} = 0. \tag{H6}$$

As a direct consequence of these assumptions, we can compute an explicit formula for the relaxation map at the boundary of the feasible set.

Lemma 6.1. *Given* $\Phi(G(x), H(x); t)$ *be such that for all* $i \in \mathcal{I}_\Phi(x; t)$

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

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

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

with

$$\alpha_i^G(x; t) = F_{Gi}(x; t) - \left.\frac{\partial \psi(x; t)}{\partial x}\right|_{x=H_i(x)} F_{Hi}(x; t),$$
$$\alpha_i^H(x; t) = F_{Hi}(x; t) - \left.\frac{\partial \psi(x; t)}{\partial x}\right|_{x=G_i(x)} F_{Gi}(x; t).$$

We now give some examples of functions ψ.

Example 2. *Some examples of functions* ψ *are:*

- $\psi(x; t) = t$ *as in [20, 15, 6] or [18].*
- $\psi(x; t) = t_1\theta_{t_2}(G(x))$*, where* $\theta_{t_2} : \mathbb{R} \to]-\infty, 1]$ *are continuously differentiable non-decreasing concave function with* $\theta(0) = 0$ *and* $\lim_{t_2\to 0} \theta_{t_2}(x) = 1\ \forall x \in \mathbb{R}_{++}$ *completed in a smooth way for negative values. This method considered in [9, 10] was called butterfly relaxation in the aforementioned papers.*

It should be noted that ψ *in* F_G *and* F_H *can be different, to get an asymmetric regularization as in [10].*

Theorem 6.2. *Assume that the relaxation map* Φ *satisfies all the above assumption with the construction in* (8)*. Then,*

$$\lim_{\|t\|\to 0} \mathcal{F}(t) = \mathcal{X},$$

where the limit is assumed pointwise.

The proof is straightforward, using that $\lim_{\|t\|\to 0} \bar{t}(t) = 0$ and the combination of (H3) with (8).

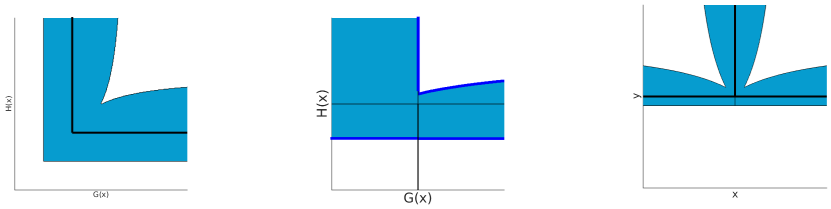


Table 1: Feasible sets of the butterfly relaxation for the three families of degenerate constraints.

6.2 Convergence Properties of the Regularization Methods

The previous section introduces a large class of regularization methods for (1). The algorithmic scheme compute a sequence of approximate stationary points for each value of a sequence of parameters $\{t_k\}$ decreasing to zero.

The main difficulty here is that convergence property are deteriorated by the approximate computation of the stationary points. It has been shown in [21, Theorem 9 and 12] or [9, Theorem 4.3] for the KDB, L-shape and butterfly relaxations that under this definition, sequences of epsilon-stationary points only converge to weak-stationary point without additional hypothesis.

In order to attain our goal to compute an M-stationary point with a realistic method, we introduce a specific definition of approximate stationary point. The definition of epsilon-stationary point called strong epsilon-stationary point, which is more stringent regarding the complementarity constraint.

In the sequel, we prove the strong convergence property of the regularization methods with this notion (Theorem 6.4) and give an algorithm (Algorithm 1).

Definition 6.3. *x^k is a strong epsilon-stationary point for $(R_t(x))$ with $\epsilon_k \geq 0$ if there exists $\eta^k \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{3q}$ such that*

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

and

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

We use here the convention that $|\eta_i^G \bar{t}_i^G| \leq \epsilon$ for $\bar{t}_i^G = \infty \Longrightarrow \eta_i^\Phi = 0$.

The representation of $\nabla\Phi$ gives that

$$\nabla \mathcal{L}^1_{R_t}(x, \eta; t) = \nabla f(x) + \sum_{i=1}^{p} \eta_i^g \nabla g_i(x) - \sum_{i=1}^{q} \nu_i^G \nabla G_i(x) - \sum_{i=1}^{q} \nu_i^H \nabla H_i(x), \tag{10}$$

where $(\eta^g, \eta^G, \eta^H, \eta^\Phi) \geq 0$ and $\nu^g := \eta^g$,

$$\nu_i^G := \begin{cases} \eta_i^G - \eta_i^\Phi \alpha_i^H(x;t), \text{ if } i \in \mathcal{I}_{CC}, \\ -\eta_i^\Phi \alpha_i^H(x;t), \text{ if } i \in \mathcal{I}_{VC}, \\ -(\eta_i^\Phi - \eta_{i+q3}^\Phi)\alpha_i^H(x;t), \text{ if } i \in \mathcal{I}_{KC}, \end{cases} \quad \text{and}$$
$$\nu_i^H := \begin{cases} \eta_i^H - \eta_i^\Phi \alpha_i^G(x;t), \text{ if } i \in \mathcal{I}_{CC}, \\ \eta_i^H - \eta_i^\Phi \alpha_i^G(x;t), \text{ if } i \in \mathcal{I}_{VC}, \\ \eta_i^H - (\eta_i^\Phi - \eta_{i+q3}^\Phi)\alpha_i^G(x;t), \text{ if } i \in \mathcal{I}_{KC}. \end{cases} \tag{11}$$

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

Theorem 6.4. *Given $\{t_k\}$ a sequence of parameters and $\{\epsilon_k\}$ a sequence of non-negative parameters such that both sequences decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \eta^k\}$ be a sequence of strong epsilon-stationary points of $(R_t(x))$ according to definition 6.3 for all $k \in \mathbb{N}$ with $x^k \to x^*$ such that G-CRSC holds at x^*. Then, x^* is an M-stationary point of (1).*

Proof. The proof relies on the use of Corollary 1. So, it is sufficient to check that up to some sufficiently large k the sequence $\{x^k\}$ and the sequence $\{\nu^k\}$ defined in (11) satisfy Assumption 5.1. In particular, we choose a sequence ν^k that satisfies condition (iv) of Assumption 5.1. This is always possible according to Remark 3.

We denote $\|\nu^k\|_\infty := \|\nu^{g,k}, \nu^{G,k}, \nu^{H,k}\|_\infty$ that without loss of generality we assume to be non-zero for k large.

- By definition of the sequence, it holds that $\left\|\nabla \mathcal{L}^1_{R_t}(x^k, \nu^k; t_k)\right\|_\infty \leq \epsilon_k$, thus as $k \to \infty$ we obtain $\nabla \mathcal{L}^1_{R_t}(x^k, \nu^k; t_k) \to 0$. A same argument gives that $\nu^{g,k} \geq 0$.

  We now show that for k sufficiently large any $\nu_i^{G,k} \leq 0$ for $i \in \mathcal{I}_{VC}^{0+}$ and $\nu_i^{H,k} \geq 0$ for $i \in \mathcal{I}_{VC}^{-0}$. Let $i \in \mathcal{I}_{VC}^{0+}$, by assumption (H2), $\alpha_i^H(x^k; t_k) \to H_i(x^*) > 0$. Thus, for k sufficiently large $\alpha_i^H(x^k; t_k) > 0$ and since $\eta_i^{\Phi,k} \geq 0$ we get $\nu_i^{G,k} \leq 0$. A straightforward adaptation of this reasoning can be used to show that $\nu_i^{H,k} \geq 0$ for $i \in \mathcal{I}_{VC}^{-0}$. So, Assumption 5.1 (i) is checked.

- It is clear by the complementarity condition in Definition 6.3 that $\nu_i^{g,k} \to 0$ for all $i \notin \mathcal{I}_g$.

  Let us now consider $i \notin \mathcal{I}_{CC}^{0+} \cup \mathcal{I}_{VC}^{0+} \cup \mathcal{I}_{KC}^{0+}$, i.e. $G_i(x^*) \neq 0$. It follows that $G_i(x^k) + \bar{t}_k \to G_i(x^*) \neq 0$. By the complementarity condition in Definition 6.3, we obtain that $\eta_i^{G,k} = 0$ for k sufficiently large.

  We now consider $\eta_i^{\Phi,k} > 0$ for k sufficiently large, otherwise the proof of this case is completed ($\nu_i^{G,k} = 0$). By Definition 6.3, it implies that $i \in \mathcal{I}_\Phi$ and so either $F_{Gi}(x^k; t_k) = 0$ or $F_{Hi}(x^k; t_k) = 0$ according to (H4). However, $F_{Gi}(x^k; t_k) = 0$ is not possible, since $F_{Gi}(x^k; t_k) \to G_i(x^*) \neq 0$ by (H5). So, $F_{Hi}(x^k; t_k) = 0$ and by (9) we have $\eta_i^{H,k} \to 0$ as ψ is non-negative and $\epsilon_k = o(\bar{t}_k)$. We can now use Lemma 6.1 to obtain

  $$\alpha_i^G(x^k; t_k) = F_{Gi}(x^k; t_k), \text{ and } \alpha_i^H(x^k; t_k) = -\left.\frac{\partial \psi(x; t_k)}{\partial x}\right|_{x = G_i(x^k)} F_{Gi}(x^k; t_k).$$

  This yields to the following inequalities for k sufficiently large

  $$\|\nu^k\|_\infty \geq \nu_i^{H,k} = \left| \frac{\nu_i^{G,k}}{\left.\frac{\partial \psi(x;t_k)}{\partial x}\right|_{x=G_i(x^k)}} \right|.$$

  It follows that $\lim_{k\to\infty} \frac{\nu_i^{G,k}}{\|\nu^k\|_\infty} = 0$.

  By the exact same way, we can prove that $\lim_{k\to\infty} \frac{\nu_i^{H,k}}{\|\nu^k\|_\infty} = 0$ for all $i \notin \mathcal{I}_{CC}^{+0} \cup \mathcal{I}_{VC}^{\pm 0} \cup \mathcal{I}_{KC}^{\pm 0}$, which concludes this part and Assumption 5.1 (ii) is checked.

- Let $i \in \mathcal{I}^{00}$. It is sufficient to consider the case $i \in \mathcal{I}_\Phi$ for k sufficiently large and in particular $\{\eta^{\Phi,k}\}$ is unbounded, otherwise this case is directly satisfied. By Definition 6.3, it implies that $i \in \mathcal{I}_\Phi$ and so either $F_{Gi}(x^k;t_k) = 0$ or $F_{Hi}(x^k;t_k) = 0$ according to (H4). We consider $F_{Hi}(x^k;t_k) = 0$ (the other case is a straightforward adaptation) and we can now use Lemma 6.1 to obtain

$$\alpha_i^G(x^k;t_k) = F_{Gi}(x^k;t_k), \text{ and } \alpha_i^H(x^k;t_k) = -\left.\frac{\partial\psi(x;t_k)}{\partial x}\right|_{x=G_i(x^k)} F_{Gi}(x^k;t_k).$$

  Since $F_{Hi}(x^k;t_k) = 0$, by (9) we have $\eta_i^{H,k} \to 0$ as ψ is non-negative and $\epsilon_k = o(\bar{t}_k)$.

  Let $0 \neq C < \infty$ such that $\lim_{k\to\infty} \nu_i^{H,k}/\|\nu\|_\infty = C$.

  If $C = 0$, the result would follow as $\lim_{k\to\infty} |\nu_i^{H,k}|/\|\nu\|_\infty \leq \infty$.

  So, assume that $C \neq 0$. Then, $\lim_{k\to\infty} -\left.\frac{\partial\psi(x;t_k)}{\partial x}\right|_{x=G_i(x^k)} \nu_i^{H,k}/\|\eta\|_\infty = 0$ by (H6). Thus, we get $\lim_{k\to\infty} \nu_i^{G,k}/\|\nu\|_\infty = \lim_{k\to\infty} \eta_i^{G,k}/\|\nu\|_\infty \geq 0$. We conclude by examining two cases: either $G_i(x^k) \geq 0$ or $G_i(x^k) \leq 0$. If $G_i(x^k) \geq 0$ the product of the limit is vanishing, since the complementarity condition and $\epsilon_k = o(\bar{t}_k)$ gives that $\lim_{k\to\infty} \nu_i^{G,k}/\|\nu\|_\infty = \lim_{k\to\infty} \eta_i^{G,k} = 0$. It is the same argument if $G_i(x^k) \leq 0$ and $\lim_{k\to\infty} \eta_i^{G,k} = 0$. Finally, if $G_i(x^k) \leq 0$ and $\lim_{k\to\infty} \nu_i^{G,k}/\|\nu\|_\infty = \lim_{k\to\infty} \eta_i^{G,k} > 0$, we get $C \geq 0$ since $G(x^k) \leq 0$ implies $F_G(x^k;t_k) \leq 0$ and $\nu_i^{H,k} = -\eta_i^{\Phi,k} F_G(x^k;t_k)$.

  This proves that Assumption 5.1 (iii) is verified.

Finally, we can apply Corollary 1 to get the result. □

Theorem (6.4) attains the ultimate goal. However it is not a trivial task to compute such a sequence of epsilon-stationary points. This is discussed in the following section.

Remark 7. *Although we do not study the existence of strong epsilon-stationary points here. The same behavior as the one explained in [26] has to be expected and the proofs can be extended. In [26, Thm 9.1 and Thm 9.2], the authors prove that reformulating* (1) *with slack variables for the degenerate constraints and then* $(R_t(x))$ *possesses strong-epsilon stationary points in the neighbourhood of an M-stationary point. It was illustrated in the aforementioned paper that this result is no longer true without slack variables.*

6.3 The Algorithm to Compute Strong Epsilon-Stationary Points

The previous section underlines the interest of considering strong epsilon-stationary points (Definition 6.3). However, it is far from obvious to see how to compute such a point in practice. We answer this essential question in this section.

6.3.1 Problem transformation

We make two transformations of $(R_t(x))$ that do not alter the theoretical properties but ease the numerical approach.

First, we add slack variables on the relaxed constraints in order to satisfy the feasibility exactly as in Definition 6.3. Consider the following non-linear parametric program $R_t(x,s)$ parametrized by t:

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

with $\lim_{\|t\|\to 0} \bar{t} = 0^+$ and the relaxation map $\Phi(s_G, s_H; t) : \mathbb{R}^{q1+q2+q3} \times \mathbb{R}^{q1+q2+q3} \to \mathbb{R}^{q1+q2+q3}$ is defined by replacing $G(x)$ and $H(x)$ by s_G and s_H in the map $\Phi(G(x), H(x); t)$.

We now penalize all the constraints that can be approximately satisfied in the Definition 6.3. The following minimization problem aims at finding $(x, s) \in \mathbb{R}^n \times \mathbb{R}^{q1+q2+q3} \times \mathbb{R}^{q1+q2+q3}$ so that

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

where ϕ is the penalty function

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

An adaptation of Theorem 6.4 gives the following result that validates the use of slack variables and the penalization approach.

Corollary 2. *Given a decreasing sequence $\{\rho_k\}$ of positive parameters and $\{\epsilon_k\}$ a sequence of non-negative parameters that decrease to zero as $k \in \mathbb{N}$ goes to infinity. Assume that $\epsilon_k = o(\bar{t}_k)$. Let $\{x^k, \eta^k\}$ be a sequence of strong epsilon-stationary points of $(P^t_\rho(x, s))$ according to definition 6.3 for all $k \in \mathbb{N}$ with $x^k \to x^*$ such that G-CRSC holds at x^*. If x^* is feasible, then it is an M-stationary point of (1).*

Proof. Assuming that x^* is feasible for (1), the result is a straightforward adaptation of Theorem 6.4. □

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

6.3.2 Active-Set Algorithm

We discuss here an active set method to solve the penalized problem $(P^t_\rho(x, s))$. This method is an extension of the method proposed in [19].

The set of points that satisfy the constraints of $(R^s_t(x, s))$ is denoted by $\mathcal{F}_{t,\bar{t}}$ and $\beta_t(x, s)$ denotes the measure of feasibility

$$\beta_t(x, s) := \phi(x, s) + \|(-s_G + \bar{t}^G)^+\|^2 + \|(-s_H + \bar{t}^H)^+\|^2 + \|(-\Phi(s_G, s_H; t))^+\|^2.$$

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

$$s_G \geq -\bar{t}^G,\ s_H \geq -\bar{t}^H,\ \Phi(s_G, s_H; t) \leq 0, \tag{12}$$

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

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

Remark 8. *It is essential to note here that active constraints act almost like bound constraints since an active constraint means that for some $i \in \{1, \ldots, q\}$ one (possibly both) of the two cases holds*

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

Considering the relaxation from Kanzow & Schwartz it is obviously a bound constraint since $\psi(s_{G,i}; t) = \psi(s_{H,i}; t) = t$. The butterfly relaxation gives $\psi(s_{G,i}; t) = t_1\theta_{t_2}(s_{H,i})$ and $\psi(s_{H,i}; t) = t_1\theta_{t_2}(s_{G,i})$. This is not

a bound constraint but we can easily use a substitution technique. This key observation is another motivation to use a formulation with slack variables.

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

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

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

$$(s_{H,i} + \alpha d_{s_{H,i}} + t_2)(s_{G,i} + \alpha d_{s_{G,i}}) - t_1(s_{H,i} + \alpha d_{s_{H,i}}) = 0. \tag{13}$$

Algorithm 1 presents an active-set scheme to solve $(P^t_\rho(x,s))$, which is described in depth in the sequel of this section. Apart from some specific parameters most of the input data are given in this algorithm through the relaxation loop that will be discussed in Algorithm 2 (page 19).

Data:
Input Data: x^{k-1}, s^{k-1}; precision $\epsilon > 0$;
$\rho_0 > 0$ initial value of ρ, ρ_{min} lower bound on the penalty parameter ;
Algorithm Parameters: $\sigma_\rho \in (0,1)$ update in ρ; $\tau_{vio} \in (0,1)$; sat:=true;
Initial estimate of the multiplier ν^0;
Begin ;
Set $j := 0$, $\rho := \rho_0$;
$(x^{k,0}, s^{k,0}, \mathcal{W}_0, A_0)$=Projection of (x^{k-1}, s^{k-1}) if not feasible for $(P^t_\rho(x,s))$;
while *sat* ***and*** $\left(\|\nabla \mathcal{L}^1_{R_t}(x^{k,j}, s^{k,j}, \eta^j; t_k))\|^2_\infty > \epsilon \|\eta^j\|_\infty$ ***or*** $\min(\eta^j) < 0$ ***or*** $\beta_{t_k}(x^{k,j}, s^{k,j}) > \epsilon\right)$ **do**
  Substitution of the variables that are fixed by the active constraints in $\mathcal{W}_j$;
  Compute a feasible direction d^j that lies in the subspace defined by the working set $\mathcal{W}_j$ (see (15)) and satisfies the conditions (SDD);
  Compute $\bar{\alpha}$ the maximum non-negative feasible step along d^j

$$\bar{\alpha} := \sup\{\alpha : (x^{k,j}, s^{k,j}) + \alpha d^j \in \mathcal{F}_{t,\bar{t}}\}$$

  Compute a step length $\alpha_j \leq \bar{\alpha}$ (see (13)) such that Armijo condition (16) holds ;
  if $\alpha_j = \bar{\alpha}$ **then**
    Update the working set $\rightarrow \mathcal{W}_{j+1}$ and compute A_{j+1} the matrix of gradients of active constraints
  $(x^{k,j+1}, s^{k,j+1}) = (x^{k,j}, s^{k,j}) + \alpha_j d^j$;
  j:=j+1 ;
  if $\beta_{t_k,\bar{t}_k}(x^{k,j+1}, s^{k,j+1}) \geq \max(\tau_{vio}\beta_{t_k,\bar{t}_k}(x^{k,j}, s^{k,j}), \epsilon)$ **then**
    $\rho := \max(\sigma_\rho \rho, \rho_{min})$
  else
    Determine the approximate multipliers $\eta^{j+1} = (\eta^G, \eta^H, \eta^\Phi)$ by solving

$$\eta^{j+1} \in \arg\min_{\nu \in \mathbb{R}^{|\mathcal{W}_j|}} \|A^T_{j+1}\eta - \nabla \Psi_\rho(x^{k,j}, s^{k,j})\|^2$$

    Relaxing rule : **if** $\exists\ i,\ \eta^{j+1}_i < 0$ ***and*** *(satisfy* (17) ***or*** $\alpha_j = 0$ *)* **then**
      Update of the working set $\mathcal{W}_{j+1}$ (with an anti-cycling rule) ;
  sat:=$\|d^j\| > \epsilon$
return: x^k, s^k, ρ or a decision of unboundedness.

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

At each step, the set $\mathcal{W}_j$ denotes the set of active constraints of the current iterate $s^{k,j}$. As pointed out in Remark 8, these active constraints fix some of the variables. Therefore, by replacing these fixed variables we

can rewrite the problem in a subspace of the initial domain. Thus, we consider the following minimization problem

$$\begin{aligned} \min_{(x,s)\in\mathbb{R}^n\times\mathbb{R}^{|\mathcal{S}_G|+|\mathcal{S}_H|}} & \Psi_\rho(x, s_{\mathcal{S}_G\cup\mathcal{S}_H}) \\ \text{s.t. } & s_{G,i} \geq -\bar{t} \text{ for } i \in \mathcal{S}_G,\ s_{H,i} \geq -\bar{t} \text{ for } i \in \mathcal{S}_H, \\ & \Phi_i(s_G, s_H; t) \leq 0 \text{ for } i \in \mathcal{S}_G \cup \mathcal{S}_H, \end{aligned} \tag{14}$$

where we denote

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

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

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

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

$$\nabla\Psi_\rho^{\mathcal{W}_j}(x, s_{\mathcal{S}_G\cup\mathcal{S}_H}) = J_{\bar{\mathcal{W}}_j}^T \nabla\Psi_\rho(x, s), \tag{15}$$

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

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

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

$$J_{\bar{\mathcal{W}}_j}^x = Id_n,$$

$$J_{\bar{\mathcal{W}}_j,i}^{s_G} = \begin{cases} e_i^T, \text{ for } i \in \mathcal{S}_G, \\ \frac{\partial\psi(x;t)}{\partial x}\Big|_{x=s_H} e_i^T, \text{ for } i \in \mathcal{I}_{GH}^{+0}, \\ 0, \text{ for } i \in (\{1,\ldots,q1\}\backslash\mathcal{S}_G)\backslash\mathcal{I}_{GH}^{+0}, \end{cases}$$

$$J_{\bar{\mathcal{W}}_j,i}^{s_H} = \begin{cases} e_i^T, \text{ for } i \in \mathcal{S}_H, \\ \frac{\partial\psi(x;t)}{\partial x}\Big|_{x=s_G} e_i^T, \text{ for } i \in \mathcal{I}_{GH}^{0+}, \\ 0, \text{ for } i \in (\{1,\ldots,q1\}\backslash\mathcal{S}_H)\backslash\mathcal{I}_{GH}^{0+}, \end{cases}$$

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

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

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

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

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

$$\Psi(z^j + \alpha_j d^j) \leq \Psi(z^j) + \tau_0 \alpha_j \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j, \ \tau_0 \in (0,1), \tag{16}$$

$$\nabla \Psi^{\mathcal{W}_j}(z^j + \alpha_j d^j)^T d^j \geq \tau_1 \nabla \Psi^{\mathcal{W}_j}(z^j)^T d^j, \ \tau_1 \in (\tau_0, 1). \tag{17}$$

If $\bar{\alpha}$ satisfies the Armijo condition (16), the active set strategy adds a new active constraint and the Wolfe condition (17) is not enforced. Otherwise, the Armijo condition requires $\alpha < \bar{\alpha}$ and the Wolfe condition is enforced.

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

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

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

6.3.3 Regularization scheme

Along this paper, we analyze an algorithm to solve (1) through a regularization scheme and an active set-penalization method to solve the sub-problems. The latter has been described in the previous sections. We now formally defined the regularization scheme in Algorithm 2.

Data: Let $z^0 = (x^0, s^0)$ be an initial point;
Let ρ_0 be an initial value of the penalty parameter;
Choose a sequence of precision $\{\epsilon_k\}$, a desired precision ϵ_∞ and a safeguard ϵ_{min};
Set $k = 0$;

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

Algorithm 2: Relaxation method for Problem (1).

As a consequence of Corollary 2, if the final point of the sequence computed by Algorithm 2 is feasible for (1) and satisfies G-CRSC, then it is an M-stationary point up to some precision. Convergence of Algorithm 1 deserves a specific treatment that has been proved in [26] in the context of complementarity constraints.

Conclusion

In this article, we focus on the study of popular classes of degenerate non-linear programs including MPCC, MPVC and the new Optimization Models with Kink Constraints, which generalize the OMCC. These problems are frequently encountered in many applications including several from optimal control as motivated here. We give theoretical results on optimality conditions, constraint qualification with their algorithmic application.

A generalized framework presented in this article is used to analyze relaxation methods that aim to converge to M-stationary points. Motivated by the approximate resolution of the sub-problems we defined a new notion of approximate stationary point. We prove existence of such approximate point in the neighborhood of an M-stationary point and provide an algorithmic strategy to compute such point.

Further research concern the implementation of this algorithmic strategy in the language Julia and its application in real-world problems.

References

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

[2] Sebastian Albrecht and Michael Ulbrich. Mathematical programs with complementarity constraints in the context of inverse optimal control for locomotion. *Optimization Methods and Software*, 32(4):670–698, 2017.

[3] John T. Betts. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. Society for Industrial and Applied Mathematics, second edition, 2010.

[4] Martin Branda, Max Bucher, Michal Červinka, and Alexandra Schwartz. Convergence of a scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. *Computational Optimization and Applications*, Feb 2018.

[5] Oleg Burdakov, Christian Kanzow, and Alexandra Schwartz. On a reformulation of mathematical programs with cardinality constraints. In David Gao, Ning Ruan, and Wenxun Xing, editors, *Advances in Global Optimization*, pages 3–14, Cham, 2015. Springer International Publishing.

[6] Oleg P. Burdakov, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016.

[7] Michal Červinka, Christian Kanzow, and Alexandra Schwartz. Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Mathematical Programming*, 160(1):353–377, Nov 2016.

[8] G. Colombo, R. Henrion, D. Hoang Nguyen, and B.S. Mordukhovich. Optimal control of the sweeping process over polyhedral controlled sets. *Journal of Differential Equations*, 260(4):3397 – 3447, 2016.

[9] Jean-Pierre Dussault, Mounir Haddou, and Tangi Migot. The new butterfly relaxation methods for mathematical program with complementarity constraint. *Optimization-Online.org*, 2016.

[10] Jean-Pierre Dussault, Mounir Haddou, and Tangi Migot. Mathematical programs with vanishing constraints: Constraint qualifications, their applications and a new regularization method. *HAL*, 2018.

[11] Mingbin Feng, John E. Mitchell, Jong-Shi Pang, Xin Shen, and Andreas Wächter. Complementarity formulations of l0-norm optimization problems. *Industrial Engineering and Management Sciences. Technical Report. Northwestern University, Evanston, IL, USA*, 2013.

[12] 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.

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

[14] Lei Guo and Gui-Hua Lin. Notes on some constraint qualifications for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 156(3):600–616, 2013.

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

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[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] Christian Kirches, Andreas Potschka, Hans Georg Bock, and Sebastian Sager. A parametric active set method for quadratic programs with vanishing constraints. *Technical Report*, 2012.

[23] Jean-Claude Latombe. *Robot motion planning*. Kluwer Academic Publishers, Norwell, MA, 1991.

[24] Christian Meyer and Oliver Thoma. A priori finite element error analysis for optimal control of the obstacle problem. *SIAM Journal on Numerical Analysis*, 51(1):605–628, 2013.

[25] F. Mignot and J. P. Puel. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, 22(3):466–476, 1984.

[26] Tangi Migot, Jean-Pierre Dussault, Mounir Haddou, and Abdeslam Kadrani. How to compute a local minimum of the MPCC. *Optimization-Online.org*, 2017.

[27] Jufeng Peng, Mihai Anitescu, and Srinivas Akella. Optimal control of multiple robot systems with friction using mpcc. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 5, pages 5224–5231. IEEE, 2004.

[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] 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.

[31] Alexandre Vieira, Bernard Brogliato, and Christophe Prieur. Quadratic optimal control of linear complementarity systems: First order necessary conditions and numerical analysis. *HAL*, 2018.

[32] Gerd Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM Journal on Optimization*, 24(4):1914–1932, 2014.