Optimisation Problems with Sparsity Terms: Theory and Algorithms

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der
Julius-Maximilians-Universität Würzburg

vorgelegt von

Andreas Budi Raharja

aus Jakarta

Würzburg, 2021

1. Gutachter: Prof. Dr. Christian Kanzow
   Julius-Maximilians-Universität Würzburg
2. Gutachter: Prof. Dr. Tim Hoheisel
   McGill University

Abstract

The present thesis deals with optimisation problems with sparsity terms, either in the constraints which lead to cardinality-constrained problems or in the objective function which in turn lead to sparse optimisation problems. One of the primary aims of this work is to extend the so-called sequential optimality conditions to these two classes of problems. In recent years sequential optimality conditions have become increasingly popular in the realm of standard nonlinear programming. In contrast to the more well-known Karush-Kuhn-Tucker condition, they are genuine optimality conditions in the sense that every local minimiser satisfies these conditions without any further assumption. Lately they have also been extended to mathematical programmes with complementarity constraints. At around the same time it was also shown that optimisation problems with sparsity terms can be reformulated into problems which possess similar structures to mathematical programmes with complementarity constraints. These recent developments have become the impetus of the present work. But rather than working with the aforementioned reformulations which involve an artifical variable we shall first directly look at the problems themselves and derive sequential optimality conditions which are independent of any artificial variable. Afterwards we shall derive the weakest constraint qualifications associated with these conditions which relate them to the Karush-Kuhn-Tucker-type conditions. Another equally important aim of this work is to then consider the practicability of the derived sequential optimality conditions. The previously mentioned reformulations open up the possibilities to adapt methods which have been proven successful to handle mathematical programmes with complementarity constraints. We will show that the safeguarded augmented Lagrangian method and some regularisation methods may generate a point satisfying the derived conditions.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit Optimierungsproblemen mit dünnbesetzten Termen, und zwar entweder in der Restriktionsmenge, was zu kardinalitätsrestringierten Problemen führen, oder in der Zielfunktion, was zu Optimierungsproblemen mit dünnbesetzten Lösungen führen. Die Herleitung der sogenannten sequentiellen Optimalitätsbedingungen für diese Problemklassen ist eines der Hauptziele dieser Arbeit. Im Bereich der nichtlinearen Optimierung gibt es in jüngster Zeit immer mehr Interesse an diesen Bedingungen. Im Gegensatz zu der mehr bekannten Karush-Kuhn-Tucker Bedingung sind diese Bedingungen echte Optimalitätsbedingungen. Sie sind also in jedem lokalen Minimum ohne weitere Voraussetzung erfüllt. Vor Kurzem wurden solche Bedingungen auch für mathematische Programme mit Komplementaritätsbedingungen hergeleitet. Zum gleichen Zeitpunkt wurde es auch gezeigt, dass Optimierungsproblemen mit dünnbesetzten Termen sich als Problemen, die ähnliche Strukturen wie mathematische Programme mit Komplementaritätsbedingungen besitzen, umformulieren lassen. Diese jüngsten Entwicklungen motivieren die vorliegende Arbeit. Hier werden wir zunächst die ursprunglichen Problemen direkt betrachten anstatt mit den Umformulierungen, die eine künstliche Variable enthalten, zu arbeiten. Dies ermöglicht uns, um Optimalitätsbedingungen, die von künstlichen Variablen unabhängig sind, zu gewinnen. Danach werden wir die entsprechenden schwächsten Constraint Qualifikationen, die diese Bedingungen mit Karush-Kuhn-Tucker-ähnlichen Bedingungen verknüpfen, herleiten. Als ein weiteres Hauptziel der Arbeit werden wir dann untersuchen, ob die gerade hergeleiteten Bedingungen eine praktische Bedeutung haben. Die vor Kurzem eingeführten Umformulierungen bieten die Möglichkeiten, um die für mathematische Programme mit Komplementaritätsbedingungen gut funktionierenden Methoden hier auch anzuwenden. Wir werden zeigen, dass das safeguarded augmented Lagrangian Method und einige Regularisierungsmethoden theoretisch in der Lage sind, um einen Punkt, der den hergeleiteten Bedingungen genügt, zu generieren.

Acknowledgements

First and foremost I would like to express my gratitude to my supervisor Christian Kanzow who has given me the opportunity to pursue research in this exciting area of mathematics and from whom I first learned about optimisation. Above all I am indebted to Christian and his family's hospitality while I was staying at their house. Next, I would like to thank Alexandra Schwartz whose pioneering work together with Christian and Oleg P. Burdakov forms the backbone of this thesis and with whom I have written a couple of papers. Many thanks also go to my colleagues at the institute of mathematics, in particular Silke Korbl who has helped me overcome various bureaucratic hurdles. Last but not least I wish to thank Tim Hoheisel who has agreed to co-referee this thesis.

Contents

Chapter 1

Introduction

Let $n, m, p, s \in \mathbb{N}$, $\rho > 0$, $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, and $h \in C^1(\mathbb{R}^n, \mathbb{R}^p)$. We consider optimisation problems of the form

$$\min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0,\ h(x) = 0,\ \|x\|_0 \leq s \tag{1.1}$$

and of the form

$$\min_x f(x) + \rho\|x\|_0 \quad \text{s.t.} \quad g(x) \leq 0,\ h(x) = 0, \tag{1.2}$$

where the mapping $x \mapsto \|x\|_0$ denotes the number of nonzero components of a given vector $x \in \mathbb{R}^n$. These problems are respectively known in the literature as *cardinality constrained optimisation problems*, CC for short, and as *sparse optimisation problems*, SP for short. Throughout this thesis we assume that $s < n$ since the cardinality constraint would be redundant otherwise.

Motivated by the desire to obtain sparse solutions in a number of application areas such as portfolio optimisation [13, 14, 18] and statistical regression [13, 28], these two classes of problems have received an increasing amount of interests in recent years. Unfortunately, the presence of the mapping $\|\cdot\|_0$, which, in spite of the notation used here, does not define a norm and is not even continuous, makes these problems difficult to solve.

One way to attack (1.1) is to reformulate them as mixed-integer problems. This reformulation is the backbone of many algorithms which employ ideas from discrete optimisation, see for example [13, 14, 45, 48, 55, 56]. Nevertheless, even testing the feasibility of (1.1) is known to be NP-complete [13].

On the other hand, a popular way to tackle (1.2) is to replace $\|\cdot\|_0$ by the sparsity promoting l_1-norm $\|\cdot\|_1$ which is obviously convex and continuous. This leads us to the following problem

$$\min_x f(x) + \rho\|x\|_1 \quad \text{s.t.} \quad g(x) \leq 0,\ h(x) = 0. \tag{1.3}$$

However, a glaring problem with such approach is of course that the solution set of (1.3) may not coincide with that of (1.2). As an example let us consider the following problem

$$\min_{x\in\mathbb{R}} \left(x - \frac{1}{2}\right)^2 + \frac{1}{5}\|x\|_0. \tag{1.4}$$

It is easy to see that $\frac{1}{2}$ is the only global minimiser of this problem. Furthermore, this problem also admits a local minimiser in 0. Now let us take a look at the corresponding l_1-minimisation problem

$$\min_{x\in\mathbb{R}} \left(x - \frac{1}{2}\right)^2 + \frac{1}{5}|x|. \tag{1.5}$$

This problem is clearly convex and hence, every local minimiser is also a global minimiser. Now it is easy to verify that $\frac{2}{5}$ is the only solution of (1.5). On the other hand, as we have already shown, $\frac{2}{5}$ is not even a local minimiser of (1.4). This illustrates the need to search for another approach to solve (1.2).

A new approach to overcome the difficulty posed by the mapping $\|\cdot\|_0$ was introduced very recently in [21], see also [30] for a similar approach. To simplify the notation we define

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ h(x) = 0\}. \tag{1.6}$$

Throughout this thesis we shall assume that $X \neq \emptyset$. Now suppose that $x \in \mathbb{R}^n$. We define

$$I_\pm(x) := \{i \in \{1,\dots,n\} \mid x_i \neq 0\} \quad \text{and} \quad I_0(x) := \{i \in \{1,\dots,n\} \mid x_i = 0\}.$$

Clearly we have $\{1,\dots,n\} = I_\pm(x) \dot\cup I_0(x)$, where $\dot\cup$ denotes union of two disjoint sets. Observe that

$$\|x\|_0 = \operatorname{card}(I_\pm(x)) = \sum_{i\in I_\pm(x)} 1 = \sum_{i\in I_\pm(x)} (1-0) + \sum_{i\in I_0(x)} (1-1).$$

Thus, by defining $y \in \mathbb{R}^n$ such that

$$y_i := \begin{cases} 0 & \text{if } i \in I_\pm(x), \\ 1 & \text{if } i \in I_0(x) \end{cases}$$

we obtain

$$\|x\|_0 = \sum_{i\in I_\pm(x)} (1-y_i) + \sum_{i\in I_0(x)} (1-y_i) = \sum_{i=1}^{n}(1-y_i) = n - e^T y,$$

where $e := (1,\dots,1)^T \in \mathbb{R}^n$. This leads to the following mixed-integer reformulations

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X,\ n - e^T y \leq s,\ y \in \{0,1\}^n,\ x \circ y = 0 \tag{1.7}$$

for (1.1) and

$$\min_{x,y} f(x) + \rho\left(n - e^T y\right) \quad \text{s.t.} \quad x \in X,\ y \in \{0,1\}^n,\ x \circ y = 0 \tag{1.8}$$

for (1.2), where $\circ$ denotes the Hadamard product. By further relaxing the binary constraint $y \in \{0,1\}^n$ we obtain

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X,\ n - e^T y \leq s,\ y \leq e,\ x \circ y = 0 \tag{1.9}$$

for (1.1) as well as

$$\min_{x,y} f(x) + \rho\left(n - e^T y\right) \quad \text{s.t.} \quad x \in X,\ y \leq e,\ x \circ y = 0. \tag{1.10}$$

for (1.2). Note that (1.9) slightly differs from the relaxed reformulation in [21] since we drop the constraint $y \geq 0$ here which leads to a larger feasible set. Nevertheless, it is easy to see that all results obtained in Section 3 of [21] are applicable to our reformulation here as well. Similarly, (1.10) also slightly differs from the half complementarity reformulation of (1.2) considered in [30] as we drop the constraint $y \geq 0$.

The relaxed reformulations above can be seen as special cases of *mathematical programmes with switching constraints* [47], MPSC for short, which in turn are closely related to *mathematical programmes with complementarity constraints*, MPCC for short. Hence, it is tempting to simply apply the results known for MPSC and MPCC to derive optimality conditions for (1.9) and (1.10). However, the thus derived conditions will then depend on the auxiliary variable y. Moreover, as noted in [23], some of the results known for MPCC are not readily applicable to (1.9).

In this thesis we shall first consider (1.1) and (1.2) directly and derive first order optimality conditions for the problems. Our derivation is based on the exterior penalty technique, cf. [12, 15]. In order to handle the problems numerically, we will then turn our attention to the relaxed reformulations (1.9) and (1.10). Let us now elaborate more on the outline and the contributions of this thesis. In Section 3.1 we shall derive two first order sequential optimality conditions for (1.1). These conditions are motivated by their nonlinear programming counterparts. Subsequently we will introduce the weakest strict constraint qualifications associated with them. We shall then consider algorithms applied to

(1.9) which can theoretically generate a point satisfying at least one of the aforementioned sequential optimality conditions, namely the augmented Lagrangian method [15], the Kanzow-Schwartz regularisation method [40], and the Steffensen-Ulbrich regularisation method [54]. Afterwards we shall turn our attention to (1.2). We will first establish the relationships between (1.1) and (1.2). Then we will derive two first order sequential optimality conditions for (1.2) in Section 4.1. Subsequently we will also introduce the weakest strict constraint qualifications associated with them. Afterwards we shall establish the equivalence between the minima of (1.2) and (1.10). Lastly we will consider algorithms applied to (1.10) which can theoretically generate a point satisfying at least one of the conditions introduced in Section 4.1. Some results in this thesis have found their way into the following preprints [42, 43].

Chapter 2

Background Material

In this chapter we shall gather some relevant tools from the theory of smooth constrained optimisation and from variational analysis. For a more comprehensive treatment of these two subjects we refer the readers to [15, 29, 33, 52]. Note that throughout this thesis we shall denote by e_i the i-th unit vector and by $\mathbb{R}_+ := [0, \infty)$.

Consider the standard nonlinear optimisation problems, NLP for short,

$$\min_x \; f(x) \quad \text{s.t.} \quad x \in X. \tag{2.1}$$

For a feasible point $\hat{x} \in X$ we define

$$I_g(\hat{x}) := \{ i \in \{1, \dots, m\} \mid g_i(\hat{x}) = 0 \} .$$

Definition 2.1. *Let $\hat{x} \in X$. We say that $\hat{x}$ is a* complementary approximately Karush-Kuhn-Tucker (CAKKT) point *iff there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) \right\} \longrightarrow 0$,

(c) $\left\{ \sum_{i=1}^m \left| \lambda_i^k g_i(x^k) \right| + \sum_{i=1}^p \left| \mu_i^k h_i(x^k) \right| \right\} \longrightarrow 0$.

The CAKKT condition for (2.1) was introduced in [5]. The following theorem asserts that this condition is a genuine first order necessary optimality condition for (2.1).

Theorem 2.2 ([5, Theorem 3.3])**.** *Let $\hat{x} \in X$ be a local minimiser of* (2.1). *Then $\hat{x}$ is a CAKKT point.*

It was shown in [5] that the CAKKT condition implies another *sequential optimality condition* whose definition we shall recall next.

Definition 2.3. *Let $\hat{x} \in X$. We say that $\hat{x}$ is an* approximately Karush-Kuhn-Tucker (AKKT) point *iff there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) \right\} \longrightarrow 0$,

(c) $\forall i \notin I_g(\hat{x}) : \; \lambda_i^k = 0 \; \forall k \in \mathbb{N}$.

The AKKT condition for (2.1) was introduced in [6, 15, 50], see also [9, 51] for similar concepts in the context of MPCC. Since the CAKKT condition implies the AKKT condition, then by Theorem 2.2 the AKKT condition is also a genuine first order necessary optimality condition for (2.1). This is in contrast to the more well-known *Karush-Kuhn-Tucker* (KKT) condition whose definition we shall recall next.

Definition 2.4. *Let $\hat{x} \in X$. We say that $\hat{x}$ is a* Karush-Kuhn-Tucker (KKT) point *iff there exist multipliers $\lambda \in \mathbb{R}^m_+$ and $\mu \in \mathbb{R}^p$ such that*

(a) $0 = \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}),$

(b) $\forall i \notin I_g(\hat{x}) \; : \; \lambda_i = 0.$

Example 2.5 ([15, page 18])**.** *Consider the following problem*

$$\min_{x \in \mathbb{R}^n} \; x_1 \quad \text{s.t.} \quad \|x\|_2^2 = 0.$$

Obviously 0 *is the only feasible point of the problem. Hence, it is the unique global minimiser. However, the KKT condition clearly does not hold at* 0*. On the other hand, by Theorem 2.2 it is a CAKKT point and therefore, also an AKKT point.*

In order for the KKT condition to be a first order necessary optimality condition for (2.1) a so-called *constraint qualification* (CQ) is needed. Let us now collect some of the known CQs for (2.1).

Definition 2.6. *Let $\hat{x} \in X$. We say that the* linear independence constraint qualification (LICQ) *holds at $\hat{x}$ iff the gradients*

$$\nabla g_i(\hat{x}) \; (i \in I_g(\hat{x})), \; \nabla h_i(\hat{x}) \; (i \in \{1, \dots, p\})$$

are linearly independent.

A weaker CQ than LICQ was introduced by Mangasarian and Fromovitz in [46].

Definition 2.7. *Let I and J be two finite index sets. A set of vectors $a^i \in \mathbb{R}^n$ for $i \in I$ and $b^i \in \mathbb{R}^n$ for $i \in J$ is called* positive-linearly dependent *iff*

$$\exists (\lambda_i \geq 0 \; (i \in I), \; \mu_i \in \mathbb{R} \; (i \in J)) \neq 0 \; : \; \sum_{i \in I} \lambda_i a^i + \sum_{i \in J} \mu_i b^i = 0.$$

Otherwise these vectors are called positive-linearly independent.

Definition 2.8. *Let $\hat{x} \in X$. We say that the* Mangasarian-Fromovitz constraint qualification (MFCQ) *holds at $\hat{x}$ iff the gradients*

$$\nabla g_i(\hat{x}) \; (i \in I_g(\hat{x})), \; \nabla h_i(\hat{x}) \; (i \in \{1, \dots, p\})$$

are positive-linearly independent.

The following condition was first introduced by Qi and Wei in [50] and has since been shown to be a CQ weaker than MFCQ in [3].

Definition 2.9. *Let $\hat{x} \in X$. We say that the* constant positive linear dependence constraint qualification (CPLD) *holds at $\hat{x}$ iff for every subsets $I_1 \subseteq I_g(\hat{x})$ and $I_2 \subseteq \{1, \dots, p\}$ such that the gradients*

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

are positive-linearly dependent in $x = \hat{x}$, they are linearly dependent for all x in a neighbourhood of $\hat{x}$.

Very recently a new CQ weaker than CPLD that is closely related to the AKKT condition was introduced in [7, 15]. This CQ was previously called the *U-condition* [15] as well as the *cone-continuity property* [7] and has since been renamed as *AKKT-regularity* [8]. To define it precisely we will need the following tool from variational analysis.

Definition 2.10. *Let $l, q \in \mathbb{N}$, $\Gamma : \mathbb{R}^l \rightrightarrows \mathbb{R}^q$ be a multifunction and $\hat{z} \in \mathbb{R}^l$. The* Painlevè-Kuratowski outer/upper limit *of* $\Gamma(z)$ *as* $z \longrightarrow \hat{z}$ *is defined as*

$$\limsup_{z \longrightarrow \hat{z}} \Gamma(z) := \{\hat{w} \in \mathbb{R}^q \mid \exists\{(z^k, w^k)\} \longrightarrow (\hat{z}, \hat{w}) \text{ with } w^k \in \Gamma(z^k)\ \forall k \in \mathbb{N}\}.$$

Now let $\hat{x} \in X$. We define for each $x \in \mathbb{R}^n$ the following cone

$$K_{\hat{x}}(x) := \left\{ \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) \,\middle|\, \begin{array}{c} (\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p, \\ \lambda_i = 0\ \forall i \notin I_g(\hat{x}) \end{array} \right\}. \tag{2.2}$$

Definition 2.11. *A feasible point $\hat{x} \in X$ for* (2.1) *is said to satisfy the* AKKT-regularity condition *iff*

$$\limsup_{x \longrightarrow \hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x}).$$

The following theorem asserts that the AKKT-regularity condition is the weakest condition on the constraints which guarantees that AKKT implies KKT, cf. [7, Theorem 3.2].

Theorem 2.12. *Let $\hat{x} \in X$. Then $\hat{x}$ is AKKT-regular iff for every continuously differentiable objective function f such that AKKT holds at $\hat{x}$, the KKT condition also holds at $\hat{x}$.*

Following [7], we then say that the AKKT-regularity condition is the weakest *strict constraint qualification* associated with the AKKT condition.

In [8], the weakest strict constraint qualification associated with the CAKKT condition was introduced. Let us now recall its definition. We define for each $x \in \mathbb{R}^n$ and for each $r \in \mathbb{R}_+$ the following cone

$$K^C((x, r)) := \left\{ \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \mu_i \nabla h_i(x) \,\middle|\, \begin{array}{c} (\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p, \\ \sum_{i=1}^{m} |\lambda_i g_i(x)| + \sum_{i=1}^{p} |\mu_i h_i(x)| \leq r \end{array} \right\}. \tag{2.3}$$

Definition 2.13. *A feasible point $\hat{x} \in X$ for* (2.1) *is said to satisfy the* CAKKT-regularity condition *iff*

$$\limsup_{(x,r) \longrightarrow (\hat{x},0)} K^C((x, r)) \subseteq K^C((\hat{x}, 0)) = K_{\hat{x}}(\hat{x}).$$

Theorem 2.14 ([8, Theorem 2])**.** *Let $\hat{x} \in X$. Then $\hat{x}$ is CAKKT-regular iff for every continuously differentiable objective function f such that CAKKT holds at $\hat{x}$, the KKT condition also holds at $\hat{x}$.*

Two of the weakest CQs for (2.1) are the *Abadie CQ* (ACQ) and the *Guignard CQ* (GCQ) which were introduced in [1] and [34] respectively. Let us recall their definitions.

Definition 2.15. *Let $A \subseteq \mathbb{R}^n$ be a nonempty set and $\hat{x} \in A$.*

- *The* polar cone *of A is defined as* $A^\circ := \left\{ y \in \mathbb{R}^n \mid y^T x \leq 0\ \forall x \in A \right\}$.
- *The* Bouligand tangent cone *to A at $\hat{x}$ is given by*

$$\mathcal{T}_A(\hat{x}) := \left\{ d \in \mathbb{R}^n \,\middle|\, \exists\{x^k\} \subseteq A, \{t_k\} \subseteq \mathbb{R}_+ : \{x^k\} \longrightarrow \hat{x}, \{t_k\} \downarrow 0, \left\{ \frac{x^k - \hat{x}}{t_k} \right\} \longrightarrow d \right\}.$$

- *The* Fréchet normal cone *to A at $\hat{x}$ is given by* $\mathcal{N}_A^F(\hat{x}) := T_A(\hat{x})^\circ$. *In some literature [29, 49], this cone is also called the* Bouligand normal cone.

Definition 2.16. *Let $\hat{x} \in X$. The* linearisation cone *of X at $\hat{x}$ is given by*

$$L_X(\hat{x}) := \left\{ d \in \mathbb{R}^n \,\middle|\, \nabla g_i(\hat{x})^T d \leq 0 \ \forall i \in I_g(\hat{x}),\ \nabla h_i(\hat{x})^T d = 0 \ \forall i \in \{1, \dots, p\} \right\}.$$

Definition 2.17. *Let $\hat{x} \in X$. We say that the*

(a) Abadie constraint qualification (ACQ) *holds at $\hat{x}$ iff $\mathcal{T}_X(\hat{x}) = L_X(\hat{x})$,*

(b) Guignard constraint qualification (GCQ) *holds at $\hat{x}$ iff $\mathcal{T}_X(\hat{x})^\circ = L_X(\hat{x})^\circ$.*

The following relation holds for the CQs introduced above, cf. [8, Figure 6]

$$\text{LICQ} \implies \text{MFCQ} \implies \text{CPLD} \implies \text{AKKT-reg.} \implies \text{CAKKT-reg.} \implies \text{ACQ} \implies \text{GCQ}.$$

Chapter 3

Cardinality Constrained Optimisation Problems

3.1 Sequential Optimality Conditions

In this chapter we shall deal with (1.1). To simplify the notation we define

$$\mathcal{S} := \{x \in \mathbb{R}^n \mid \|x\|_0 \leq s\}.$$

Observe that since $\mathcal{S}$ is a level set of the lower semicontinuous function $x \mapsto \|x\|_0$, by [29, Theorem 2.5.1] we immediately get the following result.

Lemma 3.1. *$\mathcal{S}$ is a closed set.*

Let us now consider the case where $\mathcal{S}$ is the only constraint present, i.e. we have the following problem

$$\min_x f(x) \quad \text{s.t.} \quad x \in \mathcal{S}. \tag{3.1}$$

In [11], a first order necessary optimality condition for (3.1) called *basic feasibility* (BF for short) was introduced.

Definition 3.2 ([11, Definition 2.1])**.** *Let $\hat{x} \in \mathcal{S}$. We then say that $\hat{x}$ is a* BF-vector *iff*

(a) $\|\hat{x}\|_0 < s \Longrightarrow \nabla f(\hat{x}) = 0$,

(b) $\|\hat{x}\|_0 = s \Longrightarrow \nabla_i f(\hat{x}) = 0 \ \forall i \in I_\pm(\hat{x})$.

Theorem 3.3 ([11, Theorem 2.1])**.** *Let $\hat{x} \in \mathcal{S}$ be a local minimiser of* (3.1)*. Then $\hat{x}$ is a BF-vector.*

Basic feasibility is closely related to the Fréchet normal cone of the cardinality constraint.

Lemma 3.4 ([49, Theorem 2.1])**.** *Let $\hat{x} \in \mathcal{S}$. Then*

$$\mathcal{N}^F_{\mathcal{S}}(\hat{x}) = \begin{cases} span\{e_i \mid i \in I_0(\hat{x})\} & \text{if } \|\hat{x}\|_0 = s, \\ \{0\} & \text{if } \|\hat{x}\|_0 < s. \end{cases}$$

The following proposition follows immediately from Definition 3.2 and Lemma 3.4.

Proposition 3.5. *Let $\hat{x} \in \mathcal{S}$. Then*

$$\hat{x} \text{ is a BF-vector} \iff -\nabla f(\hat{x}) \in \mathcal{N}^B_{\mathcal{S}}(\hat{x}).$$

Thus, Theorem 3.3 can be viewed as a simple corollary of [29, Remark 3.1.19] and Proposition 3.5.

If f is assumed to be convex in (3.1), we can prove the converse of Theorem 3.3. We will need the following simple observation.

Lemma 3.6. *Let $\hat{x} \in \mathbb{R}^n$. Then there exists $\epsilon > 0$ such that for each $x \in B_\epsilon(\hat{x})$ we have $I_\pm(\hat{x}) \subseteq I_\pm(x)$ and hence, $\|\hat{x}\|_0 \leq \|x\|_0$.*

Theorem 3.7. *Assume that the objective function f of* (3.1) *is convex. Let $\hat{x} \in \mathcal{S}$ be a BF-vector. If $\|\hat{x}\|_0 < s$, then $\hat{x}$ is a global minimiser of* (3.1)*. Otherwise, if $\|\hat{x}\|_0 = s$, then $\hat{x}$ is a local minimiser of* (3.1).

Proof. The convexity of f implies that for each $x \in \mathbb{R}^n$, and therefore also for each $x \in \mathcal{S}$, we have

$$f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^T (x - \hat{x}).$$

If $\|\hat{x}\|_0 < s$, we have $\nabla f(\hat{x}) = 0$ and hence, $f(x) \geq f(\hat{x})\ \forall x \in \mathcal{S}$. Now if $\|\hat{x}\|_0 = s$ instead, let $\epsilon > 0$ be as in Lemma 3.6. Then for each $x \in \mathcal{S} \cap B_\epsilon(\hat{x})$ we have $s = \|\hat{x}\|_0 \leq \|x\|_0 \leq s$ and hence, $\|x\|_0 = s = \|\hat{x}\|_0$. This implies that $I_\pm(\hat{x}) = I_\pm(x)$ and therefore, by taking complement, $I_0(\hat{x}) = I_0(x)$. Thus,

$$f(x) \geq f(\hat{x}) + \sum_{i=1}^{n} \nabla_i f(\hat{x})(x_i - \hat{x}_i) = f(\hat{x}) + \sum_{i \in I_\pm(\hat{x})} 0 \cdot (x_i - \hat{x}_i) + \sum_{i \in I_0(\hat{x})} \nabla_i f(\hat{x}) \cdot (0 - 0) = f(\hat{x}). \qquad \square$$

Let us now turn our attention to (1.1). Motivated by the CAKKT-condition for (2.1) we introduce the following definition.

Definition 3.8. *Let $\hat{x} \in X \cap \mathcal{S}$. We say that $\hat{x}$ is* CC complementary approximately M-stationary (CC-CAM-stationary) *iff there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{ \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i \right\} \longrightarrow 0,$

(c) $\left\{ \sum_{i=1}^{m} \left| \lambda_i^k g_i(x^k) \right| + \sum_{i=1}^{p} \left| \mu_i^k h_i(x^k) \right| + \sum_{i=1}^{n} \left| \gamma_i^k x_i^k \right| \right\} \longrightarrow 0.$

The following theorem asserts that CC-CAM-stationarity is a first order necessary optimality condition for (1.1).

Theorem 3.9. *Let $\hat{x} \in \mathbb{R}^n$ be a local minimiser of* (1.1)*. Then $\hat{x}$ is a CC-CAM-stationary point.*

Proof. By assumption there exists $\epsilon > 0$ such that

$$f(\hat{x}) \leq f(x) \quad \forall x \in \bar{B}_\epsilon(\hat{x}) \cap (X \cap \mathcal{S}).$$

Hence, $\hat{x}$ is the unique global minimiser of

$$\min_x \ f(x) + \frac{1}{2}\|x - \hat{x}\|_2^2 \quad \text{s.t.} \quad x \in \bar{B}_\epsilon(\hat{x}) \cap (X \cap \mathcal{S}). \tag{3.2}$$

Now let $\{\alpha_k\} \subseteq \mathbb{R}_+$ such that $\{\alpha_k\} \uparrow \infty$. We consider for each $k \in \mathbb{N}$ the partially penalised problem

$$\min_x \ f(x) + \frac{1}{2}\|x - \hat{x}\|_2^2 + \frac{\alpha_k}{2}\|(g(x)_+, h(x))\|_2^2 \quad x \in \bar{B}_\epsilon(\hat{x}) \cap \mathcal{S}. \tag{3.3}$$

Observe that the objective function in (3.3) is continuously differentiable. Furthermore, by Lemma 3.1, as an intersection between a closed and a compact set the feasible set $\bar{B}_\epsilon(\hat{x}) \cap \mathcal{S}$ is compact. Hence, by Weierstraß theorem, for each $k \in \mathbb{N}$ (3.3) admits a global minimiser $x^k \in \bar{B}_\epsilon(\hat{x}) \cap \mathcal{S}$. By the compactness of $\bar{B}_\epsilon(\hat{x}) \cap \mathcal{S}$ $\{x^k\}$ has a convergent subsequence in $\bar{B}_\epsilon(\hat{x}) \cap \mathcal{S}$. Thus, by passing to a subsequence we can assume w.l.o.g. that $\{x^k\}$ converges, i.e. $\exists \bar{x} \in \bar{B}_\epsilon(\hat{x}) \cap \mathcal{S} \; : \; \{x^k\} \longrightarrow \bar{x}$. Let us now show that $\bar{x} = \hat{x}$. Obviously for each $k \in \mathbb{N}$ $\hat{x}$ is feasible for (3.3). Thus, by the definition of x^k we have for each $k \in \mathbb{N}$

$$\begin{aligned} f(x^k) + \frac{1}{2}\left\|x^k - \hat{x}\right\|_2^2 + \frac{\alpha_k}{2}\left\|\left(g(x^k)_+, h(x^k)\right)\right\|_2^2 &\leq f(\hat{x}) + \frac{1}{2}\|\hat{x} - \hat{x}\|_2^2 + \frac{\alpha_k}{2}\|(g(\hat{x})_+, h(\hat{x}))\|_2^2 \\ &\overset{\hat{x}\in X}{=} f(\hat{x}). \end{aligned} \tag{3.4}$$

Dividing both sides by α_k and letting $k \longrightarrow \infty$ then yields

$$0 \leq \frac{1}{2}\|(g(\bar{x})_+, h(\bar{x}))\|_2^2 \leq 0$$

and hence,

$$(g(\bar{x})_+, h(\bar{x})) = 0 \quad \Longleftrightarrow \quad \bar{x} \in X.$$

This implies that $\bar{x}$ is feasible for (3.2). Now from (3.4) we also obtain that

$$f(x^k) + \frac{1}{2}\left\|x^k - \hat{x}\right\|_2^2 \leq f(\hat{x}) \quad \forall k \in \mathbb{N}.$$

Letting $k \longrightarrow \infty$ then yields

$$f(\bar{x}) + \frac{1}{2}\|\bar{x} - \hat{x}\|_2^2 \leq f(\hat{x}) = f(\hat{x}) + \frac{1}{2}\|\hat{x} - \hat{x}\|_2^2.$$

Since $\hat{x}$ is the unique global minimiser of (3.2), this implies that $\bar{x} = \hat{x}$. Thus we have $\{x^k\} \longrightarrow \hat{x}$. We can then assume w.l.o.g. that $x^k \in B_\epsilon(\hat{x})$ $\forall k \in \mathbb{N}$. Hence, by the definition of x^k this implies that for each $k \in \mathbb{N}$ x^k is a local minimiser of

$$\min_x \; f(x) + \frac{1}{2}\|x - \hat{x}\|_2^2 + \frac{\alpha_k}{2}\|(g(x)_+, h(x))\|_2^2 \qquad \text{s.t.} \qquad x \in \mathcal{S}.$$

By Theorem 3.3, x^k is then a BF-vector for each $k \in \mathbb{N}$. Now define for each $k \in \mathbb{N}$ $\gamma^k \in \mathbb{R}^n$ such that

$$\begin{aligned} -\gamma^k &:= \nabla f(x^k) + x^k - \hat{x} + \sum_{i=1}^m \alpha_k \max\{0, g_i(x^k)\}\nabla g_i(x^k) + \sum_{i=1}^p \alpha_k h_i(x^k)\nabla h_i(x^k) \\ &= \nabla f(x^k) + x^k - \hat{x} + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k), \end{aligned}$$

where

$$\begin{aligned} \lambda_i^k &:= \alpha_k \max\{0, g_i(x^k)\} && \forall i \in \{1, \ldots, m\}, \\ \mu_i^k &:= \alpha_k h_i(x^k) && \forall i \in \{1, \ldots, p\}. \end{aligned}$$

By definition we clearly have $\{\lambda^k\} \subseteq \mathbb{R}^m_+$. Next we have

$$\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k) + \gamma^k\right\} = \{\hat{x} - x^k\} \longrightarrow 0.$$

Now let $k \in \mathbb{N}$. By Definition 3.2, we have for each $i \in I_\pm(x^k)$ that $\gamma_i^k = 0$ and hence, $|\gamma_i^k x_i^k| = 0$. Moreover, we also have for each $i \in I_0(x^k)$ that $x_i^k = 0$ and therefore, $|\gamma_i^k x_i^k| = 0$. This then implies that for each $k \in \mathbb{N}$ we have

$$\sum_{i=1}^n |\gamma_i^k x_i^k| = 0.$$

Now from (3.4) we also obtain for each $k \in \mathbb{N}$ that

$$f(x^k) + \frac{\alpha_k}{2}\left\|\left(g(x^k)_+, h(x^k)\right)\right\|_2^2 \leq f(\hat{x}) \implies 0 \leq \frac{\alpha_k}{2}\left\|\left(g(x^k)_+, h(x^k)\right)\right\|_2^2 \leq f(\hat{x}) - f(x^k).$$

Since $\{f(x^k)\} \longrightarrow f(\hat{x})$, we have that $\left\{ \alpha_k \left\| \left(g(x^k)_+, h(x^k) \right) \right\|_2^2 \right\} \longrightarrow 0$. Observe that for each $k \in \mathbb{N}$ we have

$$\begin{aligned}
\alpha_k \left\| \left(g(x^k)_+, h(x^k) \right) \right\|_2^2 &= \sum_{i=1}^{m} \alpha_k \max\{0, g_i(x^k)\}^2 + \sum_{i=1}^{p} \alpha_k h_i(x^k)^2 \\
&= \sum_{i=1}^{m} |\alpha_k \max\{0, g_i(x^k)\}^2| + \sum_{i=1}^{p} |\alpha_k h_i(x^k)^2| \\
&= \sum_{i=1}^{m} |\lambda_i^k \max\{0, g_i(x^k)\}| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| \\
&= \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)|,
\end{aligned}$$

where the last equality follows from the fact that $\lambda_i^k \max\{0, g_i(x^k)\} = \lambda_i^k g_i(x^k)$. Thus,

$$\left\{ \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)\}| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k| \right\} = \left\{ \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)\}| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| \right\} \longrightarrow 0. \qquad \square$$

The converse of Theorem 3.9 is false in general as the next example shows.

Example 3.10. *Consider the following problem which is adapted from [5]*

$$\min_{x \in \mathbb{R}^3} \frac{(x_2 - 2)^2}{2} \quad s.t. \quad x_1 x_2 = 0, \ \|x\|_0 \leq 2. \tag{3.5}$$

Obviously $\hat{x} := (0, 0, 1)^T$ *is a feasible point of* (3.5) *which is not a local minimiser. On the other hand, it is a CC-CAM-stationary point. Indeed, simply define for each* $k \in \mathbb{N}$ $x^k := \hat{x}$, $\mu_k := 0$, *and* $\gamma^k := (0, 2, 0)^T$.

Nevertheless, if the cardinality constraint is active, then under some additional assumptions we can prove the reverse implication.

Theorem 3.11. *Assume that in* (1.1) *the functions* f *as well as* $g_1, \dots, g_m$ *are convex and* $h_1, \dots, h_p$ *are affine-linear. Let* $\hat{x} \in X \cap \mathcal{S}$ *such that* $\|\hat{x}\|_0 = s$. *If* $\hat{x}$ *is a CC-CAM-stationary point, then it is a local minimiser of* (1.1).

Proof. Let $\epsilon > 0$ be as in Lemma 3.6. Now let $x \in X \cap \mathcal{S} \cap B_\epsilon(\hat{x})$. We shall prove that $f(x) \geq f(\hat{x})$. Observe that since $x \in \mathcal{S} \cap B_\epsilon(\hat{x})$ we have $s = \|\hat{x}\|_0 \leq \|x\|_0 \leq s$. This implies that $\|x\|_0 = s = \|\hat{x}\|_0$. Thus we clearly have $I_\pm(x) = I_\pm(\hat{x})$, which, by taking complement, implies that $I_0(x) = I_0(\hat{x})$. Now let $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ be the corresponding CC-CAM sequences for $\hat{x}$. The assumptions made on f, g_i, and h_i imply that for each $k \in \mathbb{N}$ we have

$$\begin{aligned}
&f(x) \geq f(x^k) + \nabla f(x^k)^T (x - x^k) \\
&g_i(x) \geq g_i(x^k) + \nabla g_i(x^k)^T (x - x^k) \quad \forall i = 1, \dots, m, \\
&h_i(x) = h_i(x^k) + \nabla h_i(x^k)^T (x - x^k) \quad \forall i = 1, \dots, p.
\end{aligned}$$

Moreover, since $x \in X$, we have that $g_i(x) \leq 0$ for each $i \in \{1, \dots, m\}$ and $h_i(x) = 0$ for each $i \in \{1, \dots, p\}$. Hence, we have for each $k \in \mathbb{N}$ that $0 \geq \lambda_i^k g_i(x)$ $\forall i \in \{1, \dots, m\}$ and $0 = \mu_i^k h_i(x)$ $\forall i \in \{1, \dots, p\}$. This implies that

$$\begin{aligned}
f(x) &\geq f(x^k) + \nabla f(x^k)^T (x - x^k) \\
&\geq f(x^k) + \nabla f(x^k)^T (x - x^k) + \sum_{i=1}^{m} \lambda_i^k g_i(x) + \sum_{i=1}^{p} \mu_i^k h_i(x) + \sum_{i \in I_0(x)} \gamma_i^k x_i
\end{aligned}$$

$$
\begin{aligned}
&= f(x^k) + \nabla f(x^k)^T(x - x^k) + \sum_{i=1}^{m} \lambda_i^k g_i(x) + \sum_{i=1}^{p} \mu_i^k h_i(x) + \sum_{i\in I_0(\hat{x})} \gamma_i^k x_i \\
&\geq f(x^k) + \nabla f(x^k)^T(x - x^k) + \sum_{i=1}^{m} \lambda_i^k (g_i(x^k) + \nabla g_i(x^k)^T(x - x^k)) \\
&\quad + \sum_{i=1}^{p} \mu_i^k (h_i(x^k) + \nabla h_i(x^k)^T(x - x^k)) + \sum_{i\in I_0(\hat{x})} \gamma_i^k (x_i^k + e_i^T(x - x^k)) \\
&= f(x^k) + \left(\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i\in I_0(\hat{x})} \gamma_i^k e_i \right)^T (x - x^k) \\
&\quad + \sum_{i=1}^{m} \lambda_i^k g_i(x^k) + \sum_{i=1}^{p} \mu_i^k h_i(x^k) + \sum_{i\in I_0(\hat{x})} \gamma_i^k x_i^k. \qquad (3.6)
\end{aligned}
$$

Observe that

$$
\begin{aligned}
0 &\leq \left| \sum_{i=1}^{m} \lambda_i^k g_i(x^k) + \sum_{i=1}^{p} \mu_i^k h_i(x^k) + \sum_{i\in I_0(\hat{x})} \gamma_i^k x_i^k \right| \\
&\leq \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i\in I_0(\hat{x})} |\gamma_i^k x_i^k| \\
&\leq \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k|.
\end{aligned}
$$

Since $\left\{ \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k| \right\} \longrightarrow 0$, sandwich theorem implies that

$$
\left\{ \sum_{i=1}^{m} \lambda_i^k g_i(x^k) + \sum_{i=1}^{p} \mu_i^k h_i(x^k) + \sum_{i\in I_0(\hat{x})} \gamma_i^k x_i^k \right\} \longrightarrow 0.
$$

Now let $j \in I_\pm(\hat{x})$. We clearly have

$$
0 \leq |\gamma_j^k x_j^k| \leq \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k|.
$$

Hence, it also follows that $\{|\gamma_j^k x_j^k|\} \longrightarrow 0$. Now since $\{x_j^k\} \longrightarrow \hat{x}_j \neq 0$, this implies that $\{\gamma_j^k\} \longrightarrow 0$.

Now observe that

$$
\begin{aligned}
&\left\{ \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i\in I_0(\hat{x})} \gamma_i^k e_i \right\} \\
&= \left\{ \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i \right\} - \left\{ \sum_{i\in I_\pm(\hat{x})} \gamma_i^k e_i \right\} \longrightarrow 0.
\end{aligned}
$$

Thus, letting $k \longrightarrow \infty$ in (3.6) yields $f(x) \geq f(\hat{x})$. This completes the proof. □

The following definition is motivated by the AKKT condition for (2.1).

Definition 3.12. *Let $\hat{x} \in X \cap \mathcal{S}$. We say that $\hat{x}$ is* CC approximately M-stationary (CC-AM-stationary) *iff there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}_+^m$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \right\} \longrightarrow 0,$

(c) $\forall i \notin I_g(\hat{x}) \,:\, \lambda_i^k = 0 \;\forall k \in \mathbb{N},$

(d) $\forall i \in I_\pm(\hat{x}) \,:\, \gamma_i^k = 0 \;\forall k \in \mathbb{N}.$

Remark 3.13. *In [44], a sequential optimality condition called* AW-stationarity *was introduced. It turns out that CC-AM-stationarity is essentially equivalent to AW-stationarity, see Appendix A.*

As is the case for (2.1), CC-CAM-stationarity implies CC-AM-stationarity.

Theorem 3.14. *Let $\hat{x} \in X \cap \mathcal{S}$. If it is a CC-CAM-stationary point, then it is also CC-AM-stationary.*

Proof. Let $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ be the corresponding CC-CAM sequences for $\hat{x}$. Suppose that $j \notin I_g(\hat{x})$. Observe that

$$0 \le |\lambda_j^k g_j(x^k)| \le \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k|.$$

Thus, $\{|\lambda_j^k g_j(x^k)|\} \longrightarrow 0$ and therefore, since $\{g_j(x^k)\} \longrightarrow g_j(\hat{x}) < 0$, $\{\lambda_j^k\} \longrightarrow 0$. Now define for each $k \in \mathbb{N}$ $\hat{\lambda}^k$ such that

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{if } i \notin I_g(\hat{x}). \end{cases}$$

Obviously we have $\{\hat{\lambda}^k\} \subseteq \mathbb{R}^m_+$. Observe that

$$\begin{aligned} &\nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \\ &= \nabla f(x^k) + \sum_{i \in I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \\ &= \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \right) - \left(\sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) \right). \end{aligned}$$

Thus, it follows that

$$\left\{ \nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \right\} \longrightarrow 0.$$

Suppose now that $j \in I_\pm(\hat{x})$. Observe that

$$0 \le |\gamma_j^k x_j^k| \le \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k|.$$

Thus, $\{|\gamma_j^k x_j^k|\} \longrightarrow 0$ and therefore, since $\{x_j^k\} \longrightarrow \hat{x}_j \neq 0$, $\{\gamma_j^k\} \longrightarrow 0$. Now define for each $k \in \mathbb{N}$ $\hat{\gamma}^k$ such that

$$\hat{\gamma}_i^k := \begin{cases} \gamma_i^k & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

Then we have

$$\nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k e_i$$

$$= \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \gamma_i^k e_i$$

$$= \left(\nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i \right) - \left(\sum_{i \in I_\pm(\hat{x})} \gamma_i^k e_i \right)$$

and therefore,

$$\left\{ \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \hat{\gamma}_i^k e_i \right\} \longrightarrow 0. \qquad \square$$

A simple corollary of Theorem 3.9 and Theorem 3.14 is the following.

Theorem 3.15. *Let $\hat{x} \in \mathbb{R}^n$ be a local minimiser of* (1.1). *Then $\hat{x}$ is a CC-AM-stationary point.*

By Example 3.10 and Theorem 3.14 we know that the converse of Theorem 3.15 is false in general. The next example shows that the converse of Theorem 3.14 is false in general as well.

Example 3.16. *Let us revisit* (3.5). *Every feasible point of* (3.5) *satisfies CC-AM-stationarity. Indeed, let $\hat{x} \in \mathbb{R}^3$ be feasible for* (3.5). *If $\hat{x}_2 = 0$, we can simply define for each $k \in \mathbb{N}$ $x^k := \hat{x}$, $\mu_k := 0$, and $\gamma^k := (0, 2, 0)^T$. Otherwise, we can define for each $k \in \mathbb{N}$ $x^k := \left(\frac{1}{k}, \hat{x}_2, \hat{x}_3\right)^T$, $\mu_k := k(2 - \hat{x}_2)$, and $\gamma^k := (-k(2 - \hat{x}_2)\hat{x}_2, 0, 0)^T$. Thus, $\hat{x}$ is a CC-AM-stationary point. On the other hand, for a feasible point $\hat{x} \in \mathbb{R}^3$ to be a CC-CAM-stationary point, we must have that $\hat{x}_2 \in \{0, 2\}$. Indeed, let $\hat{x} \in \mathbb{R}^3$ be feasible for* (3.5) *such that $\hat{x}_2 \notin \{0, 2\}$. Suppose that $\hat{x}$ is a CC-CAM-stationary point with the corresponding CC-CAM-sequences $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^3$ and $\{\mu_k\} \subseteq \mathbb{R}$. The conditions in Definition 3.8 then imply that*

(a) $\{x_2^k\} \longrightarrow \hat{x}_2$,

(b) $\{x_2^k - 2 + \mu_k x_1^k + \gamma_2^k\} \longrightarrow 0$,

(c) $\{\mu_k x_1^k x_2^k\} \longrightarrow 0$,

(d) $\{\gamma_2^k x_2^k\} \longrightarrow 0$.

(a) and (b) then imply that

$$\{(x_2^k - 2)x_2^k + \mu_k x_1^k x_2^k + \gamma_2^k x_2^k\} \longrightarrow 0.$$

On the other hand we have $\{(x_2^k - 2)x_2^k\} \longrightarrow (\hat{x}_2 - 2)\hat{x}_2 \neq 0$ since $\hat{x}_2 \notin \{0, 2\}$. Thus, by (c) and (d) we have

$$\{(x_2^k - 2)x_2^k + \mu_k x_1^k x_2^k + \gamma_2^k x_2^k\} \longrightarrow (\hat{x}_2 - 2)\hat{x}_2 \neq 0.$$

This leads to a contradiction.

In the rest of this section we shall revisit (3.1), i.e. we now assume that $X = \mathbb{R}^n$ in (1.1). In this case, CC-CAM-stationarity coincides with CC-AM-stationarity.

Theorem 3.17. *Let $\hat{x} \in \mathcal{S}$. Then*

$$\hat{x} \text{ is a CC-CAM-stationarity point} \iff \hat{x} \text{ is a CC-AM-stationary point.}$$

Proof. In light of Theorem 3.14, we only need to prove the reverse implication. Suppose that $\hat{x}$ is CC-AM-stationary and let $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^n$ be the corresponding CC-AM sequences. Now if $i \in I_\pm(\hat{x})$, then we know that $\gamma_i^k = 0 \ \forall k \in \mathbb{N}$. Thus, we have for each $k \in \mathbb{N}$

$$\sum_{i=1}^{n} |\gamma_i^k x_i^k| = \sum_{i \in I_0(\hat{x})} |\gamma_i^k x_i^k|.$$

Now since $\{\nabla f(x^k) + \gamma^k\} \longrightarrow 0$ and $\{\nabla f(x^k)\} \longrightarrow \nabla f(\hat{x})$, we then have for each $i \in I_0(\hat{x})$ that $\{\gamma_i^k\} \longrightarrow -\nabla_i f(\hat{x})$ and therefore, $\{|\gamma_i^k x_i^k|\} \longrightarrow |-\nabla_i f(\hat{x}) \cdot 0| = 0$. Hence,

$$\left\{ \sum_{i=1}^{n} |\gamma_i^k x_i^k| \right\} = \left\{ \sum_{i \in I_0(\hat{x})} |\gamma_i^k x_i^k| \right\} \longrightarrow 0. \qquad \square$$

Thus, for the rest of this section a statement holds for CC-CAM-stationarity iff it holds for CC-AM-stationarity. Let us now establish the relationship between the sequential optimality conditions introduced in this section and basic feasibility, see also [19, Theorem 3.55].

Theorem 3.18. *Let $\hat{x} \in \mathcal{S}$.*

(a) If $\hat{x}$ is a BF-vector, then it is also CC-AM-stationary.

(b) If $\hat{x}$ is a CC-AM-stationary point with $\|\hat{x}\|_0 = s$, then it is a BF-vector. Thus, if the cardinality constraint is active, then CC-AM-stationarity coincides with basic feasibility.

Proof. (a) Simply define for each $k \in \mathbb{N}$ $x^k := \hat{x}$ and $y^k := -\nabla f(\hat{x}) = -\nabla f(x^k)$. Note that by Definition 3.2, we have for each $i \in I_\pm(\hat{x})$ that $y_i^k = -\nabla_i f(\hat{x}) = 0$ $\forall k \in \mathbb{N}$.
(b) By Definition 3.12 there exist sequences $\{x^k\}, \{y^k\} \subseteq \mathbb{R}^n$ such that

(b_1) $\{x^k\} \longrightarrow \hat{x}$, (b_2) $\{\nabla f(x^k) + y^k\} \longrightarrow 0$, (b_3) $\forall i \in I_\pm(\hat{x}) : y_i^k = 0$ $\forall k \in \mathbb{N}$.

Now let $i \in I_\pm(\hat{x})$. By (b_2) and (b_3), we have $\{\nabla_i f(x^k)\} = \{\nabla_i f(x^k) + y_i^k\} \longrightarrow 0$. On the other hand, by (b_1) we have $\{\nabla_i f(x^k)\} \longrightarrow \nabla_i f(\hat{x})$. Consequently we have $\nabla_i f(\hat{x}) = 0$. Since, by assumption $\|\hat{x}\|_0 = s$, $\hat{x}$ is then a BF-vector. □

The following example shows that for a feasible point $\hat{x}$ of (3.1), if $\|\hat{x}\|_0 < s$, then CC-AM-stationarity does not imply basic feasibility in general.

Example 3.19. *Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $f(x) := x_1$ and consider*

$$\min_x f(x) \quad s.t. \quad \|x\|_0 \leq 1.$$

Obviously 0 is feasible for the problem. Since $\nabla f(0) = (1,0)^T \neq 0$, it is not a BF-vector. On the other hand, the feasibility of 0 immediately implies that it is also CC-AM-stationary, see Remark 3.37.

3.2 Sequential Constraint Qualifications

Let us now return to (1.1). We define for each $x \in \mathbb{R}^n$ and each $r \in \mathbb{R}_+$

$$K^C((x,r)) := \left\{ \begin{array}{c|c} y + \sum_{i=1}^m \lambda_i \nabla g_i(x) & (\lambda, \mu, y) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n \\ + \sum_{i=1}^p \mu_i \nabla h_i(x) & \sum_{i=1}^m |\lambda_i g_i(x)| + \sum_{i=1}^p |\mu_i h_i(x)| + \sum_{i=1}^n |y_i x_i| \leq r \end{array} \right\}. \tag{3.7}$$

Now we can translate Definition 3.8 into the language of variational analysis.

Theorem 3.20. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$\hat{x} \text{ is a CC-CAM-stationary point} \iff -\nabla f(\hat{x}) \in \limsup_{(x,r)\longrightarrow(\hat{x},0)} K^C((x,r)).$$

Proof. "$\Longrightarrow$": By assumption, there exist sequences $\{x^k\}, \{y^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that the conditions in Definition 3.8 hold. Define for each $k \in \mathbb{N}$

$$u^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n y_i^k e_i$$

and

$$r_k := \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k|.$$

Then we have $\{u^k\} \longrightarrow 0$ and $\{r_k\} \longrightarrow 0$. Now define for each $k \in \mathbb{N}$

$$w^k := u^k - \nabla f(x^k) = \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i.$$

Then $\{w^k\} \longrightarrow -\nabla f(\hat{x})$. Moreover, by the definition of r_k, we clearly have for each $k \in \mathbb{N}$ that $w^k \in K^C((x^k, r_k))$. Since $\{((x^k, r_k), w^k)\} \longrightarrow ((\hat{x}, 0), -\nabla f(\hat{x}))$, we can conclude that

$$-\nabla f(\hat{x}) \in \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)).$$

"$\Longleftarrow$": By Definition 2.10, we know that there exist sequences $\{x^k\}, \{w^k\} \subseteq \mathbb{R}^n$ as well as $\{r_k\} \subseteq \mathbb{R}_+$ such that $\{((x^k, r_k), w^k)\} \longrightarrow ((\hat{x}, 0), -\nabla f(\hat{x}))$ and $w^k \in K^C((x^k, r_k))$ for each $k \in \mathbb{N}$. Hence, there exists for each $k \in \mathbb{N}$ a triple $(\lambda^k, \mu^k, \gamma^k) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that

$$w^k = \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i$$

and

$$0 \leq \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k| \leq r_k.$$

Then clearly we have

$$\left\{ \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i \right\} = \{\nabla f(x^k) + w^k\} \longrightarrow 0.$$

Since $\{r_k\} \longrightarrow 0$, it also follows that

$$\left\{ \sum_{i=1}^{m} |\lambda_i^k g_i(x^k)| + \sum_{i=1}^{p} |\mu_i^k h_i(x^k)| + \sum_{i=1}^{n} |\gamma_i^k x_i^k| \right\} \longrightarrow 0. \qquad \square$$

Let us now recall the CC-M-stationary concept introduced in [21].

Definition 3.21. *Let $\hat{x} \in X \cap \mathcal{S}$. We then say that $\hat{x}$ is CC-M-stationary iff there exist multipliers $\lambda \in \mathbb{R}^m_+$, $\mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that*

(a) $0 = \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \gamma_i e_i,$

(b) $\forall i \notin I_g(\hat{x}) : \ \lambda_i = 0,$

(c) $\forall i \in I_\pm(\hat{x}) : \ \gamma_i = 0.$

We then obtain the following translation.

Theorem 3.22. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$\hat{x} \text{ is a CC-M-stationary point} \iff -\nabla f(\hat{x}) \in K^C((\hat{x}, 0)).$$

Proof. "$\Longrightarrow$": Let $\lambda \in \mathbb{R}^m_+$, $\mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ be the corresponding multipliers for $\hat{x}$. Then we have

$$-\nabla f(\hat{x}) = \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \gamma_i e_i.$$

By the feasibility of $\hat{x}$ we have for each $i \in \{1, \dots, p\}$ that $h_i(\hat{x}) = 0$ and hence, $|\mu_i h_i(\hat{x})| = 0$. Now we now that $\forall i \notin I_g(\hat{x}) : \lambda_i = 0$ and hence, $|\lambda_i g_i(\hat{x})| = 0$. Suppose now that $i \in I_g(\hat{x})$. Then we have $g_i(\hat{x}) = 0$ and hence, $|\lambda_i g_i(\hat{x})| = 0$. Now let $i \in I_\pm(\hat{x})$. Since $\gamma_i = 0$ we then have $|\gamma_i \hat{x}_i| = 0$. Moreover, for each $i \in I_0(\hat{x})$ we have $\hat{x}_i = 0$ and therefore, $|\gamma_i \hat{x}_i| = 0$. Thus,

$$\sum_{i=1}^{m} |\lambda_i g_i(\hat{x})| + \sum_{i=1}^{p} |\mu_i h_i(\hat{x})| + \sum_{i=1}^{n} |\gamma_i \hat{x}_i| = 0.$$

"$\Longleftarrow$": By (3.7) there exists a triple $(\lambda, \mu, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that

$$-\nabla f(\hat{x}) = \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \gamma_i e_i$$

and

$$\sum_{i=1}^{m} |\lambda_i g_i(\hat{x})| + \sum_{i=1}^{p} |\mu_i h_i(\hat{x})| + \sum_{i=1}^{n} |\gamma_i \hat{x}_i| \leq 0.$$

Suppose that $i \notin I_g(\hat{x})$. Then since $g_i(\hat{x}) < 0$ and $|\lambda_i g_i(\hat{x})| = 0$, it follows that $\lambda_i = 0$. Similarly, for each $i \in I_\pm(\hat{x})$, since $\hat{x}_i \neq 0$ and $|\gamma_i \hat{x}_i| = 0$, we obtain that $\gamma_i = 0$. Thus, $\hat{x}$ is a CC-M-stationary point. ☐

Let us now investigate the relationship between CC-CAM- and CC-M-stationarity.

Theorem 3.23. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$K^C((\hat{x}, 0)) \subseteq \limsup_{(x,r) \to (\hat{x},0)} K^C((x, r)).$$

Proof. Let $\hat{w} \in K^C((\hat{x}, 0))$. Define for each $k \in \mathbb{N}$ $(x^k, r_k) := (\hat{x}, 0)$ and $w^k := \hat{w}$. This implies that $\{((x^k, r_k), w^k)\} \longrightarrow ((\hat{x}, 0), \hat{w})$ with $w^k = \hat{w} \in K^C((\hat{x}, 0)) = K^C((x^k, r_k))$ for each $k \in \mathbb{N}$. ☐

Corollary 3.24. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$\hat{x} \text{ is CC-M-stationary} \Longrightarrow \hat{x} \text{ is CC-CAM-stationary.}$$

Proof. Since $\hat{x}$ is CC-M-stationary, by Theorem 3.22 we have $-\nabla f(\hat{x}) \in K^C((\hat{x}, 0))$. Therefore, by Theorem 3.23, $-\nabla f(\hat{x}) \in \limsup_{(x,r) \to (\hat{x},0)} K^C((x, r))$. The assertion then follows from Theorem 3.20. ☐

The converse is not true in general as the following example shows.

Example 3.25 ([21, page 423])**.** *Consider*

$$\min_{x \in \mathbb{R}^2} x_1 + 10x_2 \quad s.t. \quad \left(x_1 - \frac{1}{2}\right)^2 + (x_2 - 1)^2 \leq 1, \ \|x\|_0 \leq 1. \tag{3.8}$$

Obviously $(1/2, 0)^T$ is the unique global minimiser. In particular, by Theorem 3.9, it is then also a CC-CAM-stationary point. On the other hand, we have for any $(\hat{\lambda}, \hat{\gamma}) \in \mathbb{R}_+ \times \mathbb{R}$ that

$$(1, 10)^T + \hat{\lambda}\,(2(1/2 - 1/2), 2(0 - 1))^T + \hat{\gamma}\,(0, 1)^T = (1, 10 - 2\hat{\lambda} + \hat{\gamma})^T \neq (0, 0)^T.$$

The following is clearly a sufficient condition for the converse of Corollary 3.24 to hold.

Definition 3.26. *A feasible point $\hat{x} \in \mathbb{R}^n$ of* (1.1) *is said to satisfy the CC-CAM-regularity condition iff*

$$\limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)) \subseteq K^C((\hat{x},0)).$$

Theorem 3.27. *Let $\hat{x} \in \mathbb{R}^n$ be a CC-CAM-stationary point of* (1.1) *which satisfies the CC-CAM-regularity condition. Then $\hat{x}$ is CC-M-stationary.*

Proof. This follows from Theorem 3.20, Definition 3.26, and Theorem 3.22. □

Example 3.28. *Suppose that* $0 \in \mathbb{R}^n$ *is feasible for* (1.1). *We then have*

$$K^C((0,0)) = \left\{ \gamma + \sum_{i=1}^m \lambda_i \nabla g_i(0) + \sum_{i=1}^p \mu_i \nabla h_i(0) \;\middle|\; \begin{array}{c} (\lambda,\mu,\gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n \\ \sum_{i=1}^m |\lambda_i g_i(0)| + \sum_{i=1}^p |\mu_i h_i(0)| \leq 0 \end{array} \right\}.$$

Clearly we have $K^C((0,0)) \subseteq \mathbb{R}^n$. On the other hand, for each $w \in \mathbb{R}^n$ we can simply define $(\lambda,\mu,\gamma) = (0,0,w) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ and we immediately obtain that $w \in K^C((0,0))$. Hence, we have $K^C((0,0)) = \mathbb{R}^n$. Consequently, $-\nabla f(0) \in \mathbb{R}^n = K^C((0,0))$, and therefore, by Theorem 3.22, 0 *is CC-M-stationary. Corollary 3.24 then implies that* 0 *is also a CC-CAM-stationary point. Moreover, the following inclusion also holds*

$$\limsup_{(x,r)\to(0,0)} K^C((x,r)) \subseteq \mathbb{R}^n = K^C((0,0)).$$

Thus, we conclude that 0 *satisfies CC-CAM-regularity as well.*

The next theorem states that CC-CAM-regularity is a strict constraint qualification with respect to CC-CAM-stationarity.

Theorem 3.29. *Let $\hat{x} \in X \cap S$. Suppose that for every continuously differentiable function $f \in C^1(\mathbb{R}^n, \mathbb{R})$ the following implication holds*

$$\hat{x} \text{ is CC-CAM-stationary} \implies \hat{x} \text{ is CC-M-stationary.}$$

Then $\hat{x}$ satisfies CC-CAM-regularity.

Proof. By Theorem 3.20 and Theorem 3.22 for each $f \in C^1(\mathbb{R}^n, \mathbb{R})$ the implication

$$\hat{x} \text{ is CC-CAM-stationary} \implies \hat{x} \text{ is CC-M-stationary.}$$

is equivalent to

$$-\nabla f(\hat{x}) \in \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)) \implies -\nabla f(\hat{x}) \in K^C((\hat{x},0)).$$

Suppose now that $\hat{w} \in \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r))$. In order to prove theorem, we then only need to find an $f \in C^1(\mathbb{R}^n, \mathbb{R})$ such that $-\nabla f(\hat{x}) = \hat{w}$. So we simply define $f(x) := -\hat{w}^T x$. Obviously f is continuously differentiable with $-\nabla f(\hat{x}) = \hat{w}$. Hence we have $\hat{w} = -\nabla f(\hat{x}) \in K^C((\hat{x},0))$ by the above implication. Since we can do this for every $\hat{w} \in \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r))$, we then have $\limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)) \subseteq K^C((\hat{x},0))$. This completes the proof. □

We shall now translate Definition 3.12 into the language of variational analysis as well. Let $\hat{x} \in \mathbb{R}^n$ be feasible for (1.1). Then we define for each $x \in \mathbb{R}^n$

$$K_{\hat{x}}(x) := \left\{ \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \mu_i \nabla h_i(x) + \gamma \;\middle|\; \begin{array}{c} (\lambda,\mu,\gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n, \\ \lambda_i = 0 \ \forall i \notin I_g(\hat{x}), \\ \gamma_i = 0 \ \forall i \in I_\pm(\hat{x}) \end{array} \right\}. \tag{3.9}$$

Theorem 3.30. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$\hat{x} \text{ is a CC-AM-stationary point} \iff -\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x).$$

Proof. "$\Longrightarrow$": By assumption, there exist sequences $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that the conditions in Definition 3.12 hold. Now define

$$u^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \gamma^k.$$

Then we have $\{u^k\} \longrightarrow 0$. Next we define

$$w^k := u^k - \nabla f(x^k) = \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \gamma^k.$$

Clearly we have $\{w^k\} \longrightarrow -\nabla f(\hat{x})$. Moreover, by the last two conditions in Definition 3.12 we also have $w^k \in K_{\hat{x}}(x^k)$ for each $k \in \mathbb{N}$. Hence we have $-\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x)$.

"$\Longleftarrow$": By Definition 2.10, there exist sequences $\{x^k\}, \{w^k\} \subseteq \mathbb{R}^n$ such that $\{x^k\} \longrightarrow \hat{x}$, $\{w^k\} \longrightarrow -\nabla f(\hat{x})$, and $w^k \in K_{\hat{x}}(x^k)$ for each $k \in \mathbb{N}$. Now by (3.9), for each $k \in \mathbb{N}$ there exist $(\lambda^k, \mu^k, \gamma^k) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that

- $w^k = \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \gamma^k$,
- $\forall i \notin I_g(\hat{x}) : \ \lambda_i^k = 0$,
- $\forall i \in I_\pm(\hat{x}) : \ \gamma_i^k = 0$.

Furthermore, we have

$$\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \gamma^k\right\} = \{\nabla f(x^k) + w^k\} \longrightarrow \nabla f(\hat{x}) - \nabla f(\hat{x}) = 0. \qquad \square$$

The following simple observation proves to be useful.

Lemma 3.31. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$K^C((\hat{x}, 0)) = K_{\hat{x}}(\hat{x}).$$

From Theorem 3.22 and Lemma 3.31 we immediately obtain the following.

Corollary 3.32. *Let $\hat{x} \in X \cap \mathcal{S}$. Then*

$$\hat{x} \text{ is a CC-M-stationary point} \iff -\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x}).$$

An immediate consequence of Corollary 3.24 and Theorem 3.14 is the following.

Corollary 3.33. *Let $\hat{x} \in \mathbb{R}^n$ be feasible for* (1.1). *Then*

$$\hat{x} \text{ is CC-M-stationary} \Longrightarrow \hat{x} \text{ is CC-AM-stationary}.$$

We know from Example 3.25 and Theorem 3.14 that the converse is not true in general. In light of Theorem 3.30 and Corollary 3.32 the following is clearly a sufficient condition for the converse of Corollary 3.33 to hold.

Definition 3.34. *A feasible point $\hat{x} \in \mathbb{R}^n$ of* (1.1) *is said to satisfy the CC-AM-regularity condition iff*

$$\limsup_{x \to \hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x}).$$

Theorem 3.35. *Let $\hat{x} \in \mathbb{R}^n$ be a CC-AM-stationary point of* (1.1) *which satisfies the CC-AM-regularity condition. Then $\hat{x}$ is CC-M-stationary.*

Proof. Since $\hat{x}$ is a CC-AM-stationary point, Theorem 3.30 then implies that

$$-\nabla f(\hat{x}) \in \limsup_{x \to \hat{x}} K_{\hat{x}}(x).$$

By Definition 3.34, we then have $-\nabla f(\hat{x}) \in K_{\hat{x}}(\hat{x})$ and hence, by Corollary 3.32, $\hat{x}$ is CC-M-stationary. □

The following theorem states that CC-AM-regularity is a strict constraint qualification with respect to CC-AM-stationarity. We omit the proof since it is similar to the proof of Theorem 3.29.

Theorem 3.36. *Let $\hat{x} \in X \cap S$. Suppose that for every continuously differentiable function $f \in C^1(\mathbb{R}^n, \mathbb{R})$ the following implication holds*

$$\hat{x} \text{ is CC-AM-stationary} \implies \hat{x} \text{ is CC-M-stationary.}$$

Then $\hat{x}$ satisfies CC-AM-regularity.

Remark 3.37. *Note that by the same reasoning as in Example 3.28, if $0 \in \mathbb{R}^n$ is feasible for* (1.1), *then it is a CC-AM-stationary point which satisfies CC-AM-regularity.*

Let us now establish the relationship between CC-CAM-regularity and CC-AM-regularity.

Theorem 3.38. *Let $\hat{x} \in X \cap S$. Then*

$$\limsup_{(x,r) \to (\hat{x},0)} K^C((x,r)) \subseteq \limsup_{x \to \hat{x}} K_{\hat{x}}(x).$$

Proof. Let $\hat{w} \in \limsup_{(x,r) \to (\hat{x},0)} K^C((x,r))$. Then there exist sequences $\{x^k\}, \{w^k\} \subseteq \mathbb{R}^n$ and $\{r_k\} \subseteq \mathbb{R}_+$ such that $\{((x^k, r_k), w^k)\} \to ((\hat{x}, 0), \hat{w})$ and $w^k \in K^C((x^k, r_k))$ for each $k \in \mathbb{N}$. Now let $k \in \mathbb{N}$. Since $w^k \in K^C((x^k, r_k))$, there exists a triple $(\lambda^k, \mu^k, \gamma^k) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$ such that

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \gamma^k$$

and

$$\sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k| \leq r_k.$$

Let $i \notin I_g(\hat{x})$. Since $\{g_i(x^k)\} \to g_i(\hat{x}) < 0$, we can assume w.l.o.g. that $g_i(x^k) < 0 \; \forall k \in \mathbb{N}$. Thus,

$$0 \leq |\lambda_i^k g_i(x^k)| \leq r_k \implies 0 \leq |\lambda_i^k| \leq \frac{r_k}{|g_i(x^k)|}.$$

Since $\{r_k\} \to 0$, we then have $\{\lambda_i^k\} \to 0$. Now let $i \in I_\pm(\hat{x})$. Since $\{x_i^k\} \to \hat{x}_i \neq 0$, we can assume w.l.o.g. that $x_i^k \neq 0 \forall k \in \mathbb{N}$. Thus,

$$0 \leq |\gamma_i^k x_i^k| \leq r_k \implies 0 \leq |\gamma_i^k| \leq \frac{r_k}{|x_i^k|}.$$

Thus, since $\{r_k\} \longrightarrow 0$, we also have $\{\gamma_i^k\} \longrightarrow 0$. Now define $\hat{\lambda}^k \in \mathbb{R}^m$ and $\hat{\gamma}^k \in \mathbb{R}^n$ such that

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{if } i \notin I_g(\hat{x}) \end{cases} \quad \wedge \quad \hat{\gamma}_i^k := \begin{cases} \gamma_i^k & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

Note that $\{\hat{\lambda}^k\} \subseteq \mathbb{R}_+^m$ since $\{\lambda^k\} \subseteq \mathbb{R}_+^m$. Now define

$$\begin{aligned} u^k &:= w^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k g_i(x^k) - \sum_{i \in I_\pm(\hat{x})} \gamma_i^k e_i \\ &= \sum_{i \in I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \gamma_i^k e_i \\ &= \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \hat{\gamma}_i^k e_i. \end{aligned}$$

Clearly we have $\{u^k\} \longrightarrow \hat{w}$ and $u^k \in K_{\hat{x}}(x^k)$ for each $k \in \mathbb{N}$. This then proves the desired inclusion. □

An immediate consequence of Lemma 3.31 and Theorem 3.38 is the following.

Corollary 3.39. *Let $\hat{x} \in X \cap S$. The following implication then holds*

$$\hat{x} \text{ satisfies CC-AM-regularity} \implies \hat{x} \text{ satisfies CC-CAM-regularity.}$$

Let us now investigate how CC-CAM-regularity and CC-AM-regularity relate to the other CC-tailored CQs defined in [21, 23].

Remark 3.40. *Let $\hat{x} \in X \cap S$. As remarked in [21], $\hat{x}$ is then a CC-M-stationary point iff it is a KKT point of the tightened nonlinear programme TNLP($\hat{x}$)*

$$\min_x f(x) \quad s.t. \quad g(x) \le 0, \; h(x) = 0, \; x_i = 0 \; (i \in I_0(\hat{x})). \tag{3.10}$$

Taking a closer look at Definition 3.8 and Definition 3.12, then it is also clear that $\hat{x}$ is a CC-CAM-stationary point iff $\hat{x}$ is a CAKKT-stationary point of (3.10) *and $\hat{x}$ is a CC-AM-stationary point iff $\hat{x}$ is an AKKT-stationary point of* (3.10)*. Moreover, it is also easy to see that $\hat{x}$ satisfies CC-AM-regularity iff $\hat{x}$ satisfies AKKT-regularity with respect to* (3.10)*. Likewise, $\hat{x}$ satisfies CC-CAM-regularity iff $\hat{x}$ satisfies CAKKT-regularity with respect to* (3.10).

Definition 3.41 ([23, Definition 3.11])**.** *Let $\hat{x} \in \mathbb{R}^n$ be feasible for* (1.1). *Then $\hat{x}$ satisfies*

(a) CC-LICQ *iff the gradients*

$$\nabla g_i(\hat{x}) \; (i \in I_g(\hat{x})), \; \nabla h_i(\hat{x}) \; (i \in \{1, \dots, p\}), \; e_i \; (i \in I_0(\hat{x}))$$

are linearly independent;

(b) CC-MFCQ *iff the gradients*

$$\nabla g_i(\hat{x}) \; (i \in I_g(\hat{x})), \quad \text{and} \quad \nabla h_i(\hat{x}) \; (i \in \{1, \dots, p\}), \; e_i \; (i \in I_0(\hat{x}))$$

are positive-linearly independent;

(c) CC-CPLD *iff for any subsets $I_1 \subseteq I_g(\hat{x})$, $I_2 \subseteq \{1, \dots, p\}$, and $I_3 \subseteq I_0(\hat{x})$ such that the gradients*

$$\nabla g_i(x) \; (i \in I_1), \quad \text{and} \quad \nabla h_i(x) \; (i \in I_2), \; e_i \; (i \in I_3)$$

are positive-linearly dependent in $x = \hat{x}$, they are linearly dependent in a neighbourhood of $\hat{x}$.

For the CC-tailored CQs above we have the following relation

$$\text{CC-LICQ} \implies \text{CC-MFCQ} \implies \text{CC-CPLD}.$$

Now recall from [21] that a feasible point $\hat{x}$ of (1.1) is said to satisfy any of the CC-tailored CQ above iff it satisfies the corresponding CQ for TNLP($\hat{x}$). The following implication is then immediate.

Corollary 3.42. *CC-CPLD ⟹ CC-AM-regularity ⟹ CC-CAM-regularity.*

Note that if g and h in (1.1) are affine-linear, then CC-CPLD obviously holds at every feasible point of (1.1). Thus, by Corollary 3.42 we also obtain the following.

Corollary 3.43. *If g and h in* (1.1) *are affine-linear, then every feasible point of* (1.1) *satisfies CC-AM-regularity, and therefore CC-CAM-regularity as well.*

3.3 Relaxed Reformulation

We shall now turn our attention to the relaxed reformulation (1.9) which was introduced in [21]. Let us first gather the results obtained in Section 3 of [21] which are applicable to (1.9). For proofs we refer to [21]. Note that we will denote the feasible set of (1.9) by Z.

Theorem 3.44. *[21, Theorem 3.2] Let $\hat{x} \in \mathbb{R}^n$. Then $\hat{x}$ is feasible for* (1.1) *iff there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for* (1.9).

Theorem 3.45. *[21, Theorem 3.2] Let $\hat{x} \in \mathbb{R}^n$. Then $\hat{x}$ is a global minimiser of* (1.1) *iff there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a global minimiser of* (1.9).

Theorem 3.46. *[21, Theorem 3.4] Let $\hat{x} \in \mathbb{R}^n$ be a local minimiser of* (1.1). *Then there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a local minimiser of* (1.9).

The converse of Theorem 3.46 is false in general, see [21, Example 3]. In order for the converse to hold, we need an additional assumption.

Theorem 3.47. *[21, Theorem 3.6] Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a local minimiser of* (1.9) *with $\|\hat{x}\|_0 = s$. Then $\hat{x}$ is a local minimiser of* (1.1).

Since (1.9) is an instance of (2.1), in light of Theorem 3.46, Theorem 2.2, and [15, Theorem 3.1], for every local minimiser $\hat{x} \in \mathbb{R}^n$ of (1.1) there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CAKKT and an AKKT point of (1.9). Let us now establish the relationship between CC-CAM-stationarity for (1.1) and the CAKKT condition for (1.9).

Theorem 3.48. *Let $(\hat{x}, \hat{y}) \in Z$. If $(\hat{x}, \hat{y})$ is a CAKKT point of* (1.9), *then $\hat{x}$ is a CC-CAM-stationary point of* (1.1).

Proof. By assumption there exist sequences $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, $\{\zeta_k\} \subseteq \mathbb{R}_+$, $\{\eta^k\} \subseteq \mathbb{R}^n_+$, and $\{\hat{\gamma}^k\} \subseteq \mathbb{R}^n$ such that $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$ and

$$\text{(a)} \quad \left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k y_i^k e_i \right\} \longrightarrow 0,$$

$$\text{(b)} \quad \left\{ \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + |\zeta_k(n - e^T y^k - s)| + \sum_{i=1}^n |\eta_i^k (y_i^k - 1)| + \sum_{i=1}^n |\hat{\gamma}_i^k x_i^k y_i^k| \right\} \longrightarrow 0.$$

Define for each $k \in \mathbb{N}$ $\gamma^k \in \mathbb{R}^n$ such that $\gamma_i^k := \hat{\gamma}_i^k y_i^k$ $\forall i \in \{1, \dots, n\}$. Then we have

$$\left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \right\} \longrightarrow 0.$$

Moreover, since

$$\begin{aligned} 0 &\leq \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k| \\ &\leq \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + |\zeta_k(n - e^T y^k - s)| + \sum_{i=1}^n |\eta_i^k (y_i^k - 1)| + \sum_{i=1}^n |\hat{\gamma}_i^k x_i^k y_i^k|, \end{aligned}$$

we also have

$$\left\{ \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k| \right\} \longrightarrow 0.$$ □

Theorem 3.49. *Let $\hat{x} \in X \cap \mathcal{S}$. If $\hat{x}$ is a CC-CAM-stationary point, then there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CAKKT point of* (1.9).

Proof. By assumption there exist sequences $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}_+^m$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that the conditions in Definition 3.8 hold. Observe that for each $i \in \{1, \dots, n\}$ we have

$$\{\gamma_i^k x_i^k\} \longrightarrow 0.$$

This then implies that for each $i \in I_\pm(\hat{x})$ we have

$$\{\gamma_i^k\} \longrightarrow 0.$$

Now define $\hat{y} \in \mathbb{R}^n$ such that

$$\hat{y}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then $(\hat{x}, \hat{y}) \in Z$. Next we define for each $k \in \mathbb{N}$ $\mathbb{R}^n \ni y^k := \hat{y}$, $\mathbb{R}_+ \ni \zeta_k := 0$, and $\mathbb{R}_+^n \ni \eta^k := 0$. Then

$$\begin{aligned} &\left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k y_i^k e_i \right\} \\ &= \left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \gamma_i^k e_i \right\} \\ &= \left\{ \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \right\} - \left\{ \sum_{i \in I_\pm(\hat{x})} \gamma_i^k e_i \right\} \longrightarrow 0. \end{aligned}$$

Moreover, we also have

$$\left\{ -\zeta_k e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \gamma_i^k x_i^k e_i \right\} = \left\{ \sum_{i=1}^n \gamma_i^k x_i^k e_i \right\} \longrightarrow 0$$

and

$$\begin{aligned} &\left\{ \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + |\zeta_k(n - e^T y^k - s)| + \sum_{i=1}^n |\eta_i^k (y_i^k - 1)| + \sum_{i=1}^n |\gamma_i^k x_i^k y_i^k| \right\} \\ &= \left\{ \sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i \in I_0(\hat{x})} |\gamma_i^k x_i^k| \right\} \longrightarrow 0. \end{aligned}$$ □

In [44, Theorem 4.1] it was shown that every feasible point of (A.1) satisfies the AKKT condition for (A.1). Note, however, that the AKKT condition for (1.9) differs from the AKKT condition used in [44] since we drop the constraint $y \geq 0$ which leads to the absence of the multiplier sequence associated with this constraint. Consequently, the conclusion of [44, Theorem 4.1] does not hold here as the next example shows.

Example 3.50. *Consider*

$$\min_{x,y\in\mathbb{R}^3} x_3 \quad s.t. \quad 3 - e^T y \leq 2, \ y \leq e, \ x \circ y = 0. \tag{3.11}$$

Let $a \in \mathbb{R} \setminus \{0\}$. Define $\hat{x} := (0,0,a)^T$ and $\hat{y} := (1,1,0)^T$. Then $(\hat{x},\hat{y})$ is feasible for (3.11)*. We now claim that $(\hat{x},\hat{y})$ is not an AKKT point of* (3.11)*. Indeed, suppose that it is. Let $\{(x^k,y^k)\}$, $\{\zeta_k\} \subseteq \mathbb{R}_+$, $\{\eta^k\} \subseteq \mathbb{R}^3_+$, and $\{\gamma^k\} \subseteq \mathbb{R}^3$ be the corresponding AKKT sequences. Since $3 - e^T\hat{y} = 1 < 2$ and $\hat{y}_3 = 0 < 1$, it follows that $\zeta_k = 0 = \eta^k_3$ $\forall k \in \mathbb{N}$. Now condition (b) in Definition 2.3 then implies that $\{1 + \gamma^k_3 y^k_3\} \longrightarrow 0$ and $\{\gamma^k_3 x^k_3\} \longrightarrow 0$. Since $\{x^k_3\} \longrightarrow \hat{x}_3 = a \neq 0$, it follows that $\{\gamma^k_3\} \longrightarrow 0$. But this then implies that $\{\gamma^k_3 y^k_3\} \longrightarrow 0$ and hence, $\{1 + \gamma^k_3 y^k_3\} \longrightarrow 1$. This leads to a contradiction. Note that if we include the constraint $y \geq 0$ in* (3.11)*, which turns the problem into* (A.1) *instead of* (1.9)*, then $(\hat{x},\hat{y})$ is still feasible for the problem. Hence, by [44, Theorem 4.1], in the setting of* (A.1) *the AKKT condition holds at $(\hat{x},\hat{y})$ whereas for* (1.9) *it fails as we have just shown.*

Thus, in light of [44, Theorem 4.1] and Example 3.50, the constraint $y \geq 0$ seems to have a severe impact on the behaviour of the sequential optimality conditions associated with the resulting relaxed reformulation of (1.1). This elucidates the need to look at (1.1) directly in order to derive reliable optimality conditions for (1.1) which are independent of any artificial variable and the chosen reformulation.

In contrast to Theorem 3.48, the AKKT condition for (1.9) does not imply CC-AM-stationarity as the following example which is taken from [44] shows.

Example 3.51. *[44, Example 4.2] Consider*

$$\min_{x\in\mathbb{R}^2} x_2 \quad s.t. \quad x_1^2 \leq 0, \ \|x\|_0 \leq 1 \tag{3.12}$$

and the associated relaxed reformulation

$$\min_{x,y\in\mathbb{R}^2} x_2 \quad s.t. \quad x_1^2 \leq 0, \ 1 - y_1 - y_2 \leq 0, \ y \leq e, \ x \circ y = 0. \tag{3.13}$$

Let $a > 0$. Then $\hat{x} := (0,a)^T$ is feasible for (3.12)*. The only possible value for $\hat{y} \in \mathbb{R}^2$ in order for $(\hat{x},\hat{y})$ to be feasible for* (3.13) *is $(1,0)^T$. In [44, Example 4.2] it was shown that $(\hat{x},\hat{y})$ is not an AW-stationary point. Thus, by Theorem A.4, $\hat{x}$ is not a CC-AM-stationary point. On the other hand, $(\hat{x},\hat{y})$ is an AKKT point of* (3.13)*. Indeed, we define for each $k \in \mathbb{N}$ $\mathbb{R}^2 \ni x^k := \hat{x}$, $\mathbb{R}^2 \ni y^k := (1,-1/k)^T$, $\mathbb{R}_+ \ni \lambda_k := 0$, $\mathbb{R}_+ \ni \zeta_k := ka$, $\mathbb{R}^2_+ \ni \eta^k := (ka,0)^T$, and $\mathbb{R}^2 \ni \gamma^k := (0,k)^T$. Then we have*

$$(0,1)^T + \lambda_k\,(2x^k_1,0)^T + \gamma^k_1\,(y^k_1,0)^T + \gamma^k_2\,(0,y^k_2) = (0,0)^T$$

and

$$-\zeta_k\,(1,1)^T + \eta^k_1\,(1,0)^T + \eta^k_2\,(0,1)^T + \gamma^k_1\,(x^k_1,0)^T + \gamma^k_2\,(0,x^k_2) = (0,0)^T.$$

In [23], CC-tailored ACQ and GCQ were introduced by utilising (1.9). Recall from the previous chapter that for (2.1) CAKKT-regularity implies ACQ, and therefore, GCQ. Hence, ACQ and GCQ are weaker than CAKKT-regularity. This is however not the case in the context of CC. Note that since CC-ACQ and CC-GCQ defined in [23] are tailored to (1.9) rather than (1.1), their definitions depend on the auxiliary variable y. Let us now recall their definitions.

For a given vector $x \in \mathbb{R}^n$ we define

$$I_1(x) := \{i \in \{1,\dots,n\} \mid x_i = 1\}.$$

Definition 3.52. *Let $(\hat{x}, \hat{y}) \in Z$. The CC-linearisation cone of Z at $(\hat{x}, \hat{y})$ is then given by*

$$L_Z^{CC}((\hat{x}, \hat{y})) = \left\{ (d_x, d_y) \in \mathbb{R}^n \times \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla g_i(x)^T d_x \leq 0 & \forall i \in I_g(\hat{x}), \\ \nabla h_i(x)^T d_x = 0 & \forall i = 1, \dots, p, \\ -e^T d_y \leq 0, & \text{if } n - e^T \hat{y} - s = 0 \\ e_i^T d_y \leq 0 & \forall i \in I_1(\hat{y}), \\ e_i^T d_y = 0 & \forall i \in I_\pm(\hat{x}), \\ e_i^T d_x = 0 & \forall i \in I_\pm(\hat{y}), \\ (e_i^T d_x)(e_i^T d_y) = 0 & \forall i \in I_0(\hat{x}) \cap I_0(\hat{y}) \end{array} \right\}.$$

Note that the definition above slightly differs from the one given in [23] since we drop the constraint $y \geq 0$ in our relaxed reformulation. Nevertheless, it is easy to see that the results obtained in [23] are readily applicable to our reformulation as well.

Definition 3.53. *Let $(\hat{x}, \hat{y}) \in Z$. We then say that*

(a) CC-ACQ *holds at $(\hat{x}, \hat{y})$ iff $\mathcal{T}_Z((\hat{x}, \hat{y})) = L_Z^{CC}((\hat{x}, \hat{y}))$;*

(b) CC-GCQ *holds at $(\hat{x}, \hat{y})$ iff $\mathcal{T}_Z((\hat{x}, \hat{y}))^\circ = L_Z^{CC}((\hat{x}, \hat{y}))^\circ$.*

From [23] we have the following implication

$$\text{CC-CPLD} \implies \text{CC-ACQ} \implies \text{CC-GCQ}.$$

Observe that CC-CAM-regularity does not depend on the auxiliary variable y. Now suppose that $(\hat{x}, \hat{y})$ is feasible for (1.9). By Theorem 3.44, $\hat{x}$ is then feasible for (1.1). It was shown in [23, Corollary 3.8] that CC-GCQ holds at $(\hat{x}, \hat{y})$ iff GCQ holds there. The next example shows that CC-AM-regularity, and therefore also CC-CAM-regularity, may hold at $\hat{x}$ even if CC-GCQ fails to hold at $(\hat{x}, \hat{y})$.

Example 3.54 ([21, Example 4])**.** *We consider*

$$\min_{x \in \mathbb{R}^2} \; x_1 + x_2^2 \quad \text{s.t.} \quad x_1^2 + (x_2 - 1)^2 \leq 1, \; \|x\|_0 \leq 1.$$

Then $\hat{x} := (0, 0)^T$ is the unique global minimiser of the problem. By Remark 3.37, it also satisfies CC-AM-regularity. On the other hand, if we let $\hat{y} := (0, 1)^T$, then $(\hat{x}, \hat{y})$ does not satisfy GCQ, and therefore also CC-GCQ, even though $(\hat{x}, \hat{y})$ is a global minimiser of the corresponding reformulated problem.

To close this section we would like to remark on the relationship between CC-M-stationarity and another stationarity concept introduced in [21] which is called CC-S-stationarity. Let us now recall the definition of CC-S-stationarity.

Definition 3.55. *Let $(\hat{x}, \hat{y}) \in Z$. Then $(\hat{x}, \hat{y})$ is called* CC-S-stationary *iff there exist multipliers $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that*

- $0 = \nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \mu_i \nabla h_i(\hat{x}) + \sum_{i=1}^n \gamma_i e_i,$
- $\lambda_i \geq 0, \quad \lambda_i g_i(\hat{x}) = 0 \quad \forall i = 1, \dots, m,$
- $\gamma_i = 0 \quad \forall i \in I_0(\hat{y}).$

As remarked in [21], CC-S-stationarity corresponds to the KKT condition of (1.9). Suppose now that $(\hat{x}, \hat{y}) \in Z$. By the orthogonality constraint we clearly have $I_\pm(\hat{x}) \subseteq I_0(\hat{y})$. Hence, if $(\hat{x}, \hat{y})$ is a CC-S-stationary point, then it is also CC-M-stationary. The converse is not true in general, see [21, Example 4]. However, if $(\hat{x}, \hat{y})$ is CC-M-stationary, then we can simply replace $\hat{y}$ with another auxiliary variable $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary as the next proposition shows.

Proposition 3.56. *Let $(\hat{x}, \hat{y}) \in Z$. If $(\hat{x}, \hat{y})$ is a CC-M-stationary point, then there exists $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

Proof. By Theorem 3.44 $\hat{x}$ is feasible for (1.1). Now define $\hat{z} \in \mathbb{R}^n$ such that

$$\hat{z}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then $(\hat{x}, \hat{z}) \in Z$. By assumption there exists $(\lambda, \mu, \gamma) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is CC-M-stationary. Now since $I_\pm(\hat{x}) = I_0(\hat{z})$, then by Definition 3.55 we can conclude that $(\hat{x}, \hat{z})$ is CC-S-stationary with (λ, μ, γ) from before as a corresponding multiplier. □

In [23, Theorem 4.2] it was shown that if $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ is a local minimiser of (1.9) such that CC-GCQ holds at $(\hat{x}, \hat{y})$, then it is a CC-S-stationary point. By Theorem 3.9, Theorem 3.27, and Proposition 3.56 we immediately obtain the following result for CC-CAM-regularity.

Corollary 3.57. *Let $\hat{x} \in \mathbb{R}^n$ be a local minimiser of* (1.1) *such that CC-CAM-regularity holds at $\hat{x}$. Then there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CC-S-stationary point of* (1.9).

3.4 Numerical Methods

The relaxed reformulation (1.9) opens up the possibilities of applying methods developed for (2.1) and MPCC to approximate a solution of (1.1). Of particular interests to us are the augmented Lagrangian method from [4, 15], the Kanzow-Schwartz regularisation method from [40], and the Steffensen-Ulbrich regularisation method from [54].

3.4.1 An Augmented Lagrangian Method

In this subsection we consider the applicability of an augmented Lagrangian method as described in [4, 15] for (1.9). Let us first describe the algorithm. For a given penalty parameter $\alpha > 0$ the Powell-Hestenes-Rockafellar (PHR) augmented Lagrangian for (1.9) is defined as

$$L((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) := f(x) + \alpha\pi((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha)$$

where $(\lambda, \mu, \zeta, \eta, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$ and

$$\pi((x, y), \lambda, \mu, \zeta, \eta, \gamma; \alpha) := \frac{1}{2} \left\| \begin{pmatrix} \left(g(x) + \frac{\lambda}{\alpha}\right)_+, \ h(x) + \frac{\mu}{\alpha}, \ \left(n - e^T y - s + \frac{\zeta}{\alpha}\right)_+, \\ \left(y - e + \frac{\eta}{\alpha}\right)_+, \ x \circ y + \frac{\gamma}{\alpha} \end{pmatrix} \right\|_2^2$$

is the shifted quadratic penalty term. The algorithm is then given as follows.

Algorithm 3.58 (Augmented Lagrangian Method)**.**

(S_0) *Let $\lambda_{max} > 0$, $\mu_{min} < \mu_{max}$, $\zeta_{max} > 0$, $\eta_{max} > 0$, $\gamma_{min} < \gamma_{max}$, $\tau \in (0, 1)$, $\sigma > 1$, $\bar{\lambda}^1 \in [0, \lambda_{max}]^m$, $\bar{\mu}^1 \in [\mu_{min}, \mu_{max}]^p$, $\bar{\zeta}_1 \in [0, \zeta_{max}]$, $\bar{\eta}^1 \in [0, \eta_{max}]^n$, $\bar{\gamma}^1 \in [\gamma_{min}, \gamma_{max}]^n$, $\alpha_1 > 0$, and $\{\epsilon_k\} \subseteq \mathbb{R}_+$ such that $\{\epsilon_k\} \downarrow 0$. Set $k \leftarrow 1$.*

(S_1) *Compute (x^k, y^k) as an approximate solution of*

$$\min_{x,y} \ L((x, y), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)$$

satisfying

$$\|\nabla L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)\| \le \epsilon_k. \tag{3.14}$$

(S_2) Update the approximate multipliers:

- $\lambda_i^k := \max\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\} \quad \forall i = 1, \dots, m,$
- $\mu_i^k := \alpha_k h_i(x^k) + \bar{\mu}_i^k \quad \forall i = 1, \dots, p,$
- $\zeta_k := \max\{0, \alpha_k(n - e^T y^k - s) + \bar{\zeta}_k\},$
- $\eta_i^k := \max\{0, \alpha_k(y_i^k - 1) + \bar{\eta}_i^k\} \quad \forall i = 1, \dots, n,$
- $\gamma_i^k := \alpha_k x_i^k y_i^k + \bar{\gamma}_i^k \quad \forall i = 1, \dots, n.$

(S_3) Update the penalty parameter:
Define

- $U_i^k := \min\left\{-g_i(x^k), \frac{\bar{\lambda}_i^k}{\alpha_k}\right\} \quad \forall i = 1, \dots, m,$
- $V_k := \min\left\{-(n - e^T y^k - s), \frac{\bar{\zeta}_k}{\alpha_k}\right\},$
- $W_i^k := \min\left\{-(y_i^k - 1), \frac{\bar{\eta}_i^k}{\alpha_k}\right\} \quad \forall i = 1, \dots, n.$

If $k = 1$ *or*

$$\max\left\{\begin{matrix} \|U^k\|, \ \|h(x^k)\|, \ \|V_k\|, \\ \|W^k\|, \ \|x^k \circ y^k\| \end{matrix}\right\} \leq \tau \max\left\{\begin{matrix} \|U^{k-1}\|, \ \|h(x^{k-1})\|, \ \|V_{k-1}\|, \\ \|W^{k-1}\|, \ \|x^{k-1} \circ y^{k-1}\| \end{matrix}\right\}, \tag{3.15}$$

set $\alpha_{k+1} = \alpha_k$*. Otherwise set* $\alpha_{k+1} = \sigma \alpha_k$*.*

(S_4) Update the safeguarded multipliers:
Compute $\bar{\lambda}^{k+1} \in [0, \lambda_{max}]^m$, $\bar{\mu}^{k+1} \in [\mu_{min}, \mu_{max}]^p$, $\bar{\zeta}_{k+1} \in [0, \zeta_{max}]$, $\bar{\eta}^{k+1} \in [0, \eta_{max}]^n$, $\bar{\gamma}^{k+1} \in [\gamma_{min}, \gamma_{max}]^n$.

(S_5) Set $k \longleftarrow k + 1$ *and go to* (S_1)*.*

Theorem 3.59. *Suppose that the sequence* $\{x^k\}$ *generated by Algorithm 3.58 has a limit point* $\hat{x} \in \mathbb{R}^n$*, i.e.,* $\{x^k\}$ *converges on a subsequence to* $\hat{x}$*. Then the corresponding subsequence of* $\{y^k\}$ *is bounded. In particular we can then extract a limit point* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ *of* $\{(x^k, y^k)\}$*.*

Proof. Let $\hat{x} \in \mathbb{R}^n$ be a limit point of $\{x^k\}$. By passing to a subsequence we can simplify the notation and assume w.l.o.g. that $\{x^k\} \longrightarrow \hat{x}$. Define for each $k \in \mathbb{N}$

$$B^k := \nabla_y L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) = -\zeta_k e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \gamma_i^k x_i^k e_i, \tag{3.16}$$

where the last equality follows from (S_2). By (3.14) we know that $\{B^k\} \longrightarrow 0$.
$\{y^k\}$ is bounded above:
We claim that

$$\forall i \in \{1, \dots, n\} \exists c_i \in \mathbb{R} \forall k \in \mathbb{N} : \ y_i^k \leq c_i. \tag{3.17}$$

Suppose not. Then

$$\exists j \in \{1, \dots, n\} \forall c \in \mathbb{R} \exists k \in \mathbb{N} : \ c < y_j^k.$$

We can thus construct a subsequence $\{y_j^{k_l}\}$ such that $\{y_j^{k_l}\} \longrightarrow \infty$. Consequently we can assume w.l.o.g. that $y_j^{k_l} > 1$ for each $l \in \mathbb{N}$. Now observe that by definition $\{\alpha_k\}$ is a nondecreasing sequence. In particular we then obtain for each $l \in \mathbb{N}$ that

$$0 < \alpha_1 \leq \alpha_{k_l} \tag{3.18}$$

and therefore

$$\alpha_1(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l} \leq \alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l}.$$

Since $\bar{\eta}_j^{k_l}$ is by definition a bounded sequence, we then obtain that

$$\{\alpha_1(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l}\} \longrightarrow \infty.$$

Hence, it follows that

$$\{\alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l}\} \longrightarrow \infty. \tag{3.19}$$

We can then assume w.l.o.g. that for each $l \in \mathbb{N}$ we have

$$\alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l} > 0.$$

This implies that

$$\eta_j^{k_l} = \alpha_{k_l}(y_j^{k_l} - 1) + \bar{\eta}_j^{k_l} \ \forall l \in \mathbb{N}$$

and hence, by (3.19), we have $\{\eta_j^{k_l}\} \longrightarrow \infty$. Observe that for each $l \in \mathbb{N}$ we have

$$\gamma_j^{k_l} x_j^{k_l} = (\alpha_{k_l} x_j^{k_l} y_j^{k_l} + \bar{\gamma}_j^{k_l}) x_j^{k_l} = \alpha_{k_l}(x_j^{k_l})^2 y_j^{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l} \geq \bar{\gamma}_j^{k_l} x_j^{k_l}.$$

From (3.16) we then obtain for each $l \in \mathbb{N}$ that

$$B_j^{k_l} = -\zeta_{k_l} + \eta_j^{k_l} + \gamma_j^{k_l} x_j^{k_l} \geq -\zeta_{k_l} + \eta_j^{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l}$$

which is equivalent to

$$\zeta_{k_l} \geq \eta_j^{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l} - B_j^{k_l}.$$

Since $\{B_j^{k_l}\} \longrightarrow 0$ and $\{\bar{\gamma}_j^{k_l} x_j^{k_l}\}$ is bounded we then have

$$\{\eta_j^{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l} - B_j^{k_l}\} \longrightarrow \infty.$$

Consequently we have $\{\zeta_{k_l}\} \longrightarrow \infty$. By the definition of $\{\zeta_{k_l}\}$ we can then assume w.l.o.g. that

$$\zeta_{k_l} = \alpha_{k_l}(n - e^T y^{k_l} - s) + \bar{\zeta}_{k_l} \quad \forall l \in \mathbb{N}$$

and hence,

$$\{\alpha_{k_l}(n - e^T y^{k_l} - s) + \bar{\zeta}_{k_l}\} \longrightarrow \infty.$$

Since $\{\bar{\zeta}_{k_l}\}$ is a bounded sequence we then obtain

$$\{\alpha_{k_l}(n - e^T y^{k_l} - s)\} \longrightarrow \infty.$$

We can thus assume w.l.o.g. that for each $l \in \mathbb{N}$ we have

$$\{\alpha_{k_l}(n - e^T y^{k_l} - s)\} > 0$$

and therefore, since $\alpha_{k_l} > 0$,

$$n - e^T y^{k_l} - s > 0. \tag{3.20}$$

Observe that

$$n - e^T y^{k_l} - s = n - \sum_{i=1, i \neq j}^{n} y_i^{k_l} - y_j^{k_l} - s.$$

We now claim that

$$\exists i \in \{1, \dots, n\} \setminus \{j\} \ : \ \{y_i^{k_l}\} \text{ is unbounded below.} \tag{3.21}$$

Suppose not. Then

$$\forall i \in \{1, \dots, n\} \setminus \{j\} \exists d_i \in \mathbb{R} \forall l \in \mathbb{N} \ : \ d_i \leq y_i^{k_l}.$$

We thus have for each $l \in \mathbb{N}$ that

$$-\sum_{i=1,i\neq j}^{n} y_i^{k_l} \leq -\sum_{i=1,i\neq j}^{n} d_i$$

and therefore,

$$n - \sum_{i=1,i\neq j}^{n} y_i^{k_l} - y_j^{k_l} - s \leq n - \sum_{i=1,i\neq j}^{n} d_i - y_j^{k_l} - s.$$

Since $\{y_j^{k_l}\} \longrightarrow \infty$ we then have

$$\left\{ n - \sum_{i=1,i\neq j}^{n} d_i - y_j^{k_l} - s \right\} \longrightarrow -\infty$$

and consequently

$$\left\{ n - \sum_{i=1,i\neq j}^{n} y_i^{k_l} - y_j^{k_l} - s \right\} \longrightarrow -\infty$$

which implies that for all $l \in \mathbb{N}$ large enough we have

$$n - e^T y^{k_l} - s = n - \sum_{i=1,i\neq j}^{n} y_i^{k_l} - y_j^{k_l} - s < 0.$$

But this contradicts (3.20). Hence (3.21) holds. For this index i we can then construct a subsequence $\{y^{k_{l_t}}\}$ such that $\{y^{k_{l_t}}\} \longrightarrow -\infty$. We can then assume w.l.o.g. that

$$y^{k_{l_t}} < 0 \quad \forall t \in \mathbb{N}.$$

From (3.18) we then obtain for each $t \in \mathbb{N}$ that

$$\alpha_{k_{l_t}}(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}} \leq \alpha_1(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}}.$$

Since $\{\bar{\eta}\}_i^{k_{l_t}}$ is a bounded sequence, we then have

$$\{\alpha_1(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}}\} \longrightarrow -\infty.$$

Consequently we then obtain that

$$\{\alpha_{k_{l_t}}(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}}\} \longrightarrow -\infty.$$

Therefore, we can assume w.l.o.g. that

$$\alpha_{k_{l_t}}(y_i^{k_{l_t}} - 1) + \bar{\eta}_i^{k_{l_t}} < 0 \quad \forall t \in \mathbb{N}.$$

This implies that

$$\eta_i^{k_{l_t}} = 0 \quad \forall t \in \mathbb{N}$$

and hence, we obtain from (3.16) that

$$\begin{aligned} B_i^{k_{l_t}} &= -\zeta_{k_{l_t}} + \eta_i^{k_{l_t}} + \gamma_i^{k_{l_t}} x_i^{k_{l_t}} \\ &= -\zeta_{k_{l_t}} + \gamma_i^{k_{l_t}} x_i^{k_{l_t}} \\ &= -\zeta_{k_{l_t}} + (\alpha_{k_{l_t}} x_i^{k_{l_t}} y_i^{k_{l_t}} + \bar{\gamma}_i^{k_{l_t}}) x_i^{k_{l_t}} \\ &= -\zeta_{k_{l_t}} + \alpha_{k_{l_t}} (x_i^{k_{l_t}})^2 y_i^{k_{l_t}} + \bar{\gamma}_i^{k_{l_t}} x_i^{k_{l_t}} \end{aligned}$$

$$\leq -\zeta_{k_{l_t}} + \bar{\gamma}_i^{k_{l_t}} x_i^{k_{l_t}}.$$

Since $\{\bar{\gamma}_i^{k_{l_t}} x_i^{k_{l_t}}\}$ is a bounded sequence and $\{\zeta_{k_l}\} \longrightarrow \infty$ we then have

$$\{B_i^{k_{l_t}}\} \longrightarrow -\infty$$

which leads to a contradiction. Thus, (3.17) holds, which immediately implies that $\{y^k\}$ is bounded above.
$\{y^k\}$ is bounded below:
We claim that

$$\forall i \in \{1, \ldots, n\} \exists d_i \in \mathbb{R} \forall k \in \mathbb{N} : \ d_i \leq y_i^k. \tag{3.22}$$

Suppose not. Then

$$\exists j \in \{1, \ldots, n\} \forall d \in \mathbb{R} \exists k \in \mathbb{N} : \ y_j^k < d.$$

We can thus construct a subsequence $\{y_j^{k_l}\}$ such that $\{y_j^{k_l}\} \longrightarrow -\infty$. Just like in the previous case, we can then assume w.l.o.g. that for each $l \in \mathbb{N}$ we have

$$y_j^{k_l} < 0 \quad \wedge \quad \eta_j^{k_l} = 0.$$

From (3.16) we then obtain

$$\begin{aligned} B_j^{k_l} &= -\zeta_{k_l} + \eta_j^{k_l} + \gamma_j^{k_l} x_j^{k_l} \\ &= -\zeta_{k_l} + \gamma_j^{k_l} x_j^{k_l} \\ &= -\zeta_{k_l} + (\alpha_{k_l} x_j^{k_l} y_j^{k_l} + \bar{\gamma}_j^{k_l}) x_j^{k_l} \\ &= -\zeta_{k_l} + \alpha_{k_l} (x_j^{k_l})^2 y_j^{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l} \\ &\leq -\zeta_{k_l} + \bar{\gamma}_j^{k_l} x_j^{k_l} \end{aligned}$$

which is equivalent to

$$\zeta_{k_l} \leq \bar{\gamma}_j^{k_l} x_j^{k_l} - B_j^{k_l}.$$

Since $\{\bar{\gamma}_j^{k_l} x_j^{k_l}\}$ is bounded and $\{B_j^{k_l}\} \longrightarrow 0$, the sequence $\{\bar{\gamma}_j^{k_l} x_j^{k_l} - B_j^{k_l}\}$ is bounded. This impies in particular that $\{\zeta_{k_l}\}$ is bounded above, i.e.

$$\exists r \in \mathbb{R} \forall l \in \mathbb{N} : \ \zeta_{k_l} \leq r. \tag{3.23}$$

On the other hand, we have shown that

$$\forall i \in \{1, \ldots, n\} \exists c_i \in \mathbb{R} \forall k \in \mathbb{N} : \ y_i^k \leq c_i.$$

Consequently we have for each $l \in \mathbb{N}$ that

$$-\sum_{i=1, i \neq j}^{n} c_i \leq -\sum_{i=1, i \neq j}^{n} y_i^{k_l}$$

and hence,

$$n - \sum_{i=1, i \neq j}^{n} c_i - y_j^{k_l} - s \leq n - \sum_{i=1}^{n} y_i^{k_l} - s. \tag{3.24}$$

Since $\{y_i^{k_l}\} \longrightarrow -\infty$, we then have $\left\{ n - \sum_{i=1, i \neq j}^{n} c_i - y_j^{k_l} - s \right\} \longrightarrow \infty$. We can thus assume w.l.o.g. that $n - \sum_{i=1, i \neq j}^{n} c_i - y_j^{k_l} - s > 0$ for each $l \in \mathbb{N}$. From (3.18) and (3.24) we then obtain

$$0 < \alpha_1 \left(n - \sum_{i=1, i \neq j}^{n} c_i - y_j^{k_l} - s \right) \leq \alpha_{k_l} \left(n - \sum_{i=1, i \neq j}^{n} c_i - y_j^{k_l} - s \right) \leq \alpha_{k_l} \left(n - \sum_{i=1}^{n} y_i^{k_l} - s \right)$$

and therefore,

$$\alpha_1 \left(n - \sum_{i=1, i\neq j}^{n} c_i - y_j^{k_l} - s \right) + \bar{\zeta}_{k_l} \leq \alpha_{k_l} \left(n - \sum_{i=1}^{n} y_i^{k_l} - s \right) + \bar{\zeta}_{k_l}.$$

Since $\{\bar{\zeta}_{k_l}\}$ is bounded, then $\left\{ \alpha_1 \left(n - \sum_{i=1, i\neq j}^{n} c_i - y_j^{k_l} - s \right) + \bar{\zeta}_{k_l} \right\} \longrightarrow \infty$. Thus, we have

$$\left\{ \alpha_{k_l} \left(n - \sum_{i=1}^{n} y_i^{k_l} - s \right) + \bar{\zeta}_{k_l} \right\} \longrightarrow \infty. \tag{3.25}$$

We can then assume w.l.o.g. that $\alpha_{k_l} \left(n - \sum_{i=1}^{n} y_i^{k_l} - s \right) + \bar{\zeta}_{k_l} > 0$ for each $l \in \mathbb{N}$. By the definition of ζ_{k_l} we then have

$$\zeta_{k_l} = \alpha_{k_l} \left(n - \sum_{i=1}^{n} y_i^{k_l} - s \right) + \bar{\zeta}_{k_l} \quad \forall l \in \mathbb{N}.$$

Thus, we obtain from (3.25) that $\{\zeta_{k_l}\} \longrightarrow \infty$ which contradicts (3.23). Hence, (3.22) holds which immediately implies that $\{y^k\}$ is bounded below. □

To measure the infeasibility of a point $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ for (1.9) we consider the unshifted quadratic penalty term

$$\pi_{0,1}((x, y)) := \pi((x, y), 0, 0, 0, 0, 0; 1).$$

Clearly $(\hat{x}, \hat{y})$ is feasible for (1.9) iff $\pi_{0,1}((\hat{x}, \hat{y})) = 0$. This in turn implies that $(\hat{x}, \hat{y})$ minimises $\pi_{0,1}((x, y))$. In particular we then ought to have $\nabla \pi_{0,1}((\hat{x}, \hat{y})) = 0$.

Theorem 3.60. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 3.58. Then $\nabla \pi_{0,1}((\hat{x}, \hat{y})) = 0$.*

We omit the proof since it is essentially the same as [15, Theorem 6.3].

Suppose that the sequence $\{x^k\}$ generated by Algorithm 3.58 has a limit point $\hat{x}$. Theorem 3.59 then suggests that we can extract a limit point $(\hat{x}, \hat{y})$ of the sequence $\{(x^k, y^k)\}$. Let us first consider the case where the *generalised Kurdyka-Łojasiewicz* (GKL) inequality is satisfied by $\pi_{0,1}$ at a feasible limit point $(\hat{x}, \hat{y})$ of Algorithm 3.58.

Some comments on the GKL inequality are due. The inequality was first introduced in [5]. A continuously differentiable function $F \in C^1(\mathbb{R}^n, \mathbb{R})$ is said to satisfy the GKL inequality at $\hat{x} \in \mathbb{R}^n$ iff there exist $\delta > 0$ and $\psi : B_\delta(\hat{x}) \longrightarrow \mathbb{R}$ such that $\lim_{x \to \hat{x}} \psi(x) = 0$ and for each $x \in B_\delta(\hat{x})$ we have $|F(x) - F(\hat{x})| \leq \psi(x) \|\nabla F(x)\|$. According to [5, page 3546], the GKL inequality is a relatively mild condition. For example it is satisfied at every feasible point of (2.1) provided that all constraint functions are analytic. Now if we view (1.9) as a standard NLP, then all constraints involving the auxiliary variable y are polynomial in nature and therefore analytic. Thus, if the nonlinear constraints g_i and h_i are analytic, the GKL inequality is then automatically satisfied. The following result is an immediate consequence of [5, Theorem 5.1] and Theorem 3.48.

Theorem 3.61. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 3.58 that is feasible for (1.9). Assume that $\pi_{0,1}$ satisfies the GKL inequality at $(\hat{x}, \hat{y})$, i.e., there exist $\delta > 0$ and $\psi : B_\delta((\hat{x}, \hat{y})) \longrightarrow \mathbb{R}$ such that $\lim_{(x,y) \to (\hat{x}, \hat{y})} \psi((x, y)) = 0$ and for each $(x, y) \in B_\delta((\hat{x}, \hat{y}))$ we have*

$$|\pi_{0,1}((x, y)) - \pi_{0,1}((\hat{x}, \hat{y}))| \leq \psi((x, y)) \|\nabla \pi_{0,1}((x, y))\|. \tag{3.26}$$

Then $\hat{x}$ is a CC-CAM-stationary point.

As a direct consequence of Theorem 3.61 and Theorem 3.35, we obtain the following

Corollary 3.62. *If, in addition to the assumptions in Theorem 3.61, $\hat{x}$ also satisfies CC-CAM-regularity, then $\hat{x}$ is CC-M-stationary.*

Proposition 3.56 then implies the following.

Corollary 3.63. *Under the assumptions of Corollary 3.62, there exists $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

Suppose now that we do not assume that the GKL inequality is satisfied by $\pi_{0,1}$ at a feasible limit point $(\hat{x}, \hat{y})$ of Algorithm 3.58. By [15, Theorem 6.2] $(\hat{x}, \hat{y})$ is then still an AKKT point of (1.9). Unfortunately, due to Example 3.51, in general we cannot expect $\hat{x}$ to be a CC-AM-stationary point. Moreover, as remarked in [23, Section 3], ACQ, and therefore, by [7, Theorem 4.4], AKKT regularity, are usually violated at a feasible point of (1.9). In the context of MPCC, the authors of [38] proved the global convergence of the augmented Lagrangian method towards an MPCC-C-stationarity point under MPCC-LICQ. Here we would instead employ the CC-analogue of the quasinormality CQ [12], which is weaker than CC-CPLD, to prove the global convergence of the method towards a CC-M-stationary point. Note that even though the reformulated problems can be viewed as MPCC if nonnegativity constraints $x \geq 0$ are present [23], we would not assume the presence of nonnegativity constraints here making our results applicable in the general setting. Moreover, even in the presence of nonnegativity constraints, it was shown in [23, Remark 5.7(f)] that MPCC-LICQ, which as mentioned before was used to guarantee convergence to a stationary point in [38], is often violated at points of interests for the reformulated problems. Hence, our result is not a simple corollary of [38].

Utilising (3.10) let us now introduce a CC-tailored *quasinormality* condition.

Definition 3.64. *Let $\hat{x} \in \mathbb{R}^n$ be feasible for* (1.1). *Then $\hat{x}$ satisfies the* CC-quasinormality *condition iff $\nexists(\hat{\lambda}, \hat{\mu}, \hat{\gamma}) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n \setminus \{(0,0,0)\}$ such that*

(a) $0 = \sum_{i=1}^{m} \hat{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \hat{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \hat{\gamma}_i e_i,$

(b) $\forall i \notin I_g(\hat{x}) : \ \hat{\lambda}_i = 0,$

(c) $\forall i \in I_\pm(\hat{x}) : \ \hat{\gamma}_i = 0,$

(d) $\exists \{x^k\} \subseteq \mathbb{R}^n$ *with* $\{x^k\} \longrightarrow \hat{x}$ *such that for each* $k \in \mathbb{N}$ *we have*

- $\forall i \in \{1, \dots, m\}$ *with* $\hat{\lambda}_i > 0 : \ \hat{\lambda}_i g_i(x^k) > 0,$
- $\forall i \in \{1, \dots, p\}$ *with* $\hat{\mu}_i \neq 0 : \ \hat{\mu}_i h_i(x^k) > 0,$
- $\forall i \in \{1, \dots, n\}$ *with* $\hat{\gamma}_i \neq 0 : \ \hat{\gamma}_i x_i^k > 0.$

Obviously CC-quasinormality corresponds to the quasinormality CQ of (3.10). Furthermore, by [3], CC-CPLD then implies CC-quasinormality.

Theorem 3.65. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of $\{(x^k, y^k)\}$ such that $(\hat{x}, \hat{y})$ is feasible for* (1.9) *and $\hat{x}$ satisfies CC-quasinormality. Then $(\hat{x}, \hat{y})$ is a CC-M-stationary point.*

Proof. To simplify the notation, we assume, throughout this proof, that the entire sequence $\{(x^k, y^k)\}$ converges to $(\hat{x}, \hat{y})$. For each $k \in \mathbb{N}$ we define

$$\begin{aligned} A^k &:= \nabla_x L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\zeta}_k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k y_i^k e_i, \end{aligned}$$

where the last equality follows from (S_2), and B^k be as in (3.16). By (3.14) and since $\{\epsilon_k\} \downarrow 0$ we know that $\{A^k\} \longrightarrow 0$ and $\{B^k\} \longrightarrow 0$. Observe that by (S_2) we have $\{\lambda^k\} \subseteq \mathbb{R}^m_+$. Furthermore, by (S_3) the sequence of penalty parameters $\{\alpha_k\}$ is nondecreasing. In particular we then have

$$\alpha_k \geq \alpha_1 > 0 \; \forall k \in \mathbb{N}. \tag{3.27}$$

Let us now differentiate between 2 cases.
Case 1: $\{\alpha_k\}$ is bounded.
Observe that by (S_3), the boundedness of $\{\alpha_k\}$ implies that

$$\exists K \in \mathbb{N} \forall k \geq K \; : \; \alpha_k = \alpha_K.$$

Now let us take a closer look at (S_2). The boundedness of $\{\alpha_k\}$ immediately implies that

- $\forall i \in \{1, \dots, p\} \; : \; \{\mu_i^k\}$ is bounded,
- $\forall i \in \{1, \dots, n\} \; : \; \{\gamma_i^k y_i^k\}$ is bounded.

By passing to subsequences we can assume w.l.o.g. that these sequences converge, i.e.

- $\forall i \in \{1, \dots, p\} \exists \hat{\mu}_i \; : \; \{\mu_i^k\} \longrightarrow \hat{\mu}_i$,
- $\forall i \in \{1, \dots, n\} \exists \hat{\gamma}_i \; : \; \{\gamma_i^k y_i^k\} \longrightarrow \hat{\gamma}_i$.

Suppose now that $i \in I_\pm(\hat{x})$. By the feasibility of $(\hat{x}, \hat{y})$ we have $\hat{y}_i = 0$. Since in this case we have $\{y_i^k\} \longrightarrow 0$, it follows that

$$\hat{\gamma}_i = \lim_{k \to \infty} \gamma_i^k y_i^k = \lim_{k \to \infty} \alpha_k x_i^k (y_i^k)^2 + \lim_{k \to \infty} \bar{\gamma}_i^k y_i^k = \alpha_K \cdot 0 + \lim_{k \to \infty} \bar{\gamma}_i^k y_i^k = 0.$$

Now observe that for each $i \in \{1, \dots, m\}$ we have

$$0 \leq \lambda_i^k \leq |\alpha_k g_i(x^k) + \bar{\lambda}_i^k| \; \forall k \in \mathbb{N}.$$

Thus, $\{\lambda_i^k\}$ is bounded as well and has a convergent subsequence. By passing to subsequences we can assume w.l.o.g. that

$$\forall i \in \{1, \dots, m\} \exists \hat{\lambda}_i \; : \; \{\lambda_i^k\} \longrightarrow \hat{\lambda}_i \geq 0.$$

Now the boundedness of $\{\alpha_k\}$ and (S_3) also imply that $\left\{ \left\| U^k \right\| \right\} \longrightarrow 0$. Let $i \notin I_g(\hat{x})$. By definition, $\{\bar{\lambda}^k\}$ is bounded. Thus, by (3.27) $\left\{ \frac{\bar{\lambda}_i^k}{\alpha_k} \right\}$ is bounded as well and therefore has a convergent subsequence. Assume w.l.o.g. that $\left\{ \frac{\bar{\lambda}_i^k}{\alpha_k} \right\}$ converges and denote with a_i its limit. We then have

$$0 = \lim_{k \to \infty} \left\| U_i^k \right\| = \| \min\{-g_i(\hat{x}), a_i\} \| \quad \Longrightarrow \quad \min\{-g_i(\hat{x}), a_i\} = 0.$$

Since $-g_i(\hat{x}) > 0$ we then have $a_i = 0$. This then implies that

$$\left\{ g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k} \right\} \longrightarrow g_i(\hat{x}) + a_i = g_i(\hat{x}) < 0.$$

Hence, we can assume w.l.o.g. that

$$g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k} < 0 \quad \forall k \in \mathbb{N}.$$

By (3.27) we then obtain

$$\alpha_k g_i(x^k) + \bar{\lambda}_i^k = \alpha_k \left(g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k} \right) < 0 \quad \forall k \in \mathbb{N}.$$

Thus, by (S_2) we have

$$\lambda_i^k = \max \left\{ 0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k \right\} = 0 \quad \forall k \in \mathbb{N}. \tag{3.28}$$

As its limit we then have $\hat{\lambda}_i = 0$. By the definition of A^k, letting $k \longrightarrow \infty$ then yields

$$0 = \nabla f(\hat{x}) + \sum_{i=1}^{m} \hat{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \hat{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \hat{\gamma}_i e_i$$

and we conclude that $(\hat{x}, \hat{y})$ is a CC-M-stationary point.
Case 2: $\{\alpha_k\}$ is unbounded.
Since $\{\alpha_k\}$ is nondecreasing, we have $\{\alpha_k\} \longrightarrow \infty$. Now define for each $k \in \mathbb{N}$

$$\tilde{\gamma}_i^k := \gamma_i^k y_i^k \quad \forall i \in \{1, \dots, n\}.$$

We claim that $\left\{ \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\}$ is bounded. Suppose not. By passing to a subsequence we can assume w.l.o.g. that $\left\{ \left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\| \right\} \longrightarrow \infty$. Consequently, the normed sequence $\left\{ \frac{(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \right\}$ is bounded with a constant length 1. Therefore, it has a convergent subsequence. Again, by passing to a subsequence we can assume w.l.o.g. that the whole sequence converges, i.e.

$$\exists \left(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{\zeta}, \tilde{\eta}\right) \neq 0 \ : \ \left\{ \frac{\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\|} \right\} \longrightarrow \left(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{\zeta}, \tilde{\eta}\right).$$

Observe that for each $i \in \{1, \dots, m\}$ we have

$$\frac{\lambda_i^k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\|} \geq 0 \quad \forall k \in \mathbb{N}.$$

As such

$$\tilde{\lambda}_i \geq 0 \quad \forall i \in \{1, \dots, m\} \tag{3.29}$$

Now suppose that $i \notin I_g(\hat{x})$. Then $g_i(\hat{x}) < 0$. By definition, $\left\{ \bar{\lambda}_i^k \right\}$ is a bounded sequence. Consequently we have

$$\left\{ \alpha_k g_i(x^k) + \bar{\lambda}_i^k \right\} \longrightarrow -\infty.$$

We can thus assume w.l.o.g. that

$$\alpha_k g_i(x^k) + \bar{\lambda}_i^k < 0 \quad \forall k \in \mathbb{N}$$

which immediately implies that

$$\lambda_i^k = 0 \quad \forall k \in \mathbb{N}.$$

As such we have

$$\tilde{\lambda}_i = \lim_{k \to \infty} \frac{\lambda_i^k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\|} = \lim_{k \to \infty} 0 = 0. \tag{3.30}$$

Now suppose that $i \in I_{\pm}(\hat{x})$. Since $(\hat{x}, \hat{y})$ is feasible, then we have $\hat{y}_i = 0$. Furthermore, by definition $\{\bar{\eta}_i^k\}$ is bounded. Consequently we have

$$\left\{ \alpha_k \left(y_i^k - 1\right) + \bar{\eta}_i^k \right\} \longrightarrow -\infty.$$

We can thus assume w.l.o.g. that

$$\alpha_k \left(y_i^k - 1\right) + \bar{\eta}_i^k < 0 \quad \forall k \in \mathbb{N}$$

which then implies that

$$\eta_i^k = 0 \quad \forall k \in \mathbb{N}. \tag{3.31}$$

Now we claim that $\tilde{\gamma}_i = 0$. Suppose not. Since $\left\{ \frac{\tilde{\gamma}_i^k}{\left\|(\lambda^k,\mu^k,\tilde{\gamma}^k,\zeta^k,\eta^k)\right\|} \right\} \longrightarrow \tilde{\gamma}_i$ we can then assume w.l.o.g. that $\tilde{\gamma}_i^k \neq 0 \; \forall k \in \mathbb{N}$. Since $\tilde{\gamma}_i^k = \gamma_i^k y_i^k$, this then implies that $y_i^k \neq 0 \; \forall k \in \mathbb{N}$. We then have

$$B_i^k = -\zeta_k + \eta_i^k + \gamma_i^k x_i^k \overset{(3.31)}{=} -\zeta_k + \gamma_i^k x_i^k = -\zeta_k + \frac{\tilde{\gamma}_i^k}{y_i^k} x_i^k. \tag{3.32}$$

Rearranging and dividing (3.32) by $\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|$ then gives us

$$\frac{B_i^k + \zeta_k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} = \frac{\tilde{\gamma}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \cdot x_i^k \cdot \frac{1}{y_i^k}. \tag{3.33}$$

Observe that the left hand side of (3.33) converges. On the other hand, since

$$\left\{ \frac{\tilde{\gamma}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} x_i^k \right\} \longrightarrow \tilde{\gamma}_i \hat{x}_i \neq 0$$

and $\{y_i^k\} \longrightarrow 0$, the right hand side diverges. This leads to a contradiction. Hence we have

$$\tilde{\gamma}_i = 0 \quad \forall i \in I_\pm(\hat{x}). \tag{3.34}$$

Now we claim that $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0$. Suppose not. Then since $\left(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}, \tilde{\zeta}, \tilde{\eta}\right) \neq 0$, it follows that $\left(\tilde{\zeta}, \tilde{\eta}\right) \neq 0$. Suppose now that $i \in I_0(\hat{y})$. Since $\{y_k^i\} \longrightarrow \hat{y}_i$ and $\{\bar{\eta}_i^k\}$ is a bounded sequence, we then have

$$\left\{ \alpha_k \left(y_i^k - 1\right) + \bar{\eta}_i^k \right\} \longrightarrow -\infty.$$

Just like before, we can then assume w.l.o.g. that

$$\eta_i^k = 0 \quad \forall k \in \mathbb{N} \tag{3.35}$$

and thus,

$$\tilde{\eta}_i = \lim_{k\to\infty} \frac{\eta_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} = \lim_{k\to\infty} 0 = 0.$$

Hence, we have

$$(\zeta, \eta_i \;\; (i \in I_\pm(\hat{y}))) \neq 0. \tag{3.36}$$

Now let $i \in I_\pm(\hat{y})$. Since $\hat{y}_i \neq 0$ and $\{y_i^k\} \longrightarrow \hat{y}_i$, we can assume w.l.o.g. that $y_i^k \neq 0 \; \forall k \in \mathbb{N}$. We then have for each $k \in \mathbb{N}$ that

$$B_i^k = -\zeta_k + \eta_i^k + \gamma_i^k x_i^k = -\zeta_k + \eta_i^k + \frac{\tilde{\gamma}_i^k}{y_i^k} x_i^k. \tag{3.37}$$

Rearranging and dividing (3.37) by $\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|$ then yields

$$\frac{B_i^k + \zeta_k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} = \frac{\eta_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} + \frac{\tilde{\gamma}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \cdot x_i^k \cdot \frac{1}{y_i^k}. \tag{3.38}$$

By assumption, $\tilde{y}_i = 0$. Consequently, letting $k \longrightarrow \infty$ in (3.38) yields

$$\tilde{\zeta} = \eta_i + 0 \cdot \hat{x}_i \cdot \frac{1}{\hat{y}_i} = \tilde{\eta}_i. \tag{3.39}$$

From (3.36) we then obtain

$$\tilde{\zeta} \neq 0 \quad \wedge \quad \tilde{\eta}_i = \tilde{\zeta} \neq 0 \; \forall i \in I_{\pm}(\hat{y}).$$

Observe that since by definition $\zeta_k \geq 0 \; \forall k \in \mathbb{N}$ and $\left\{ \frac{\zeta_k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \right\} \longrightarrow \tilde{\zeta}$, it follows that $\tilde{\zeta} \geq 0$. Thus we have $\tilde{\zeta} > 0$. Furthermore, we can then assume w.l.o.g. that $\zeta_k > 0 \; \forall k \in \mathbb{N}$. This implies that

$$\zeta_k = \alpha_k \left(n - e^T y^k - s \right) + \bar{\zeta}_k.$$

We then have

$$\begin{aligned} 0 < \tilde{\zeta} &= \lim_{k \to \infty} \frac{\zeta_k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} \\ &= \lim_{k \to \infty} \frac{\alpha_k \left(n - e^T y^k - s \right)}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} + \lim_{k \to \infty} \frac{\bar{\zeta}_k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} \\ &= \lim_{k \to \infty} \frac{\alpha_k \left(n - e^T y^k - s \right)}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|}, \end{aligned}$$

since $\{\bar{\zeta}_k\}$ is bounded by definition. Consequently we can assume w.l.o.g. that

$$\alpha_k \left(n - e^T y^k - s \right) > 0 \quad \forall k \in \mathbb{N}$$

and therefore, since $\alpha_k > 0$,

$$n - e^T y^k - s > 0 \quad \forall k \in \mathbb{N}. \tag{3.40}$$

By assumption $(\hat{x}, \hat{y})$ is feasible and hence, $n - e^T \hat{y} - s \leq 0$. Thus, we obtain from (3.40) that $n - e^T y^k - s > n - e^T \hat{y} - s$ and therefore,

$$e^T \hat{y} > e^T y^k \quad \forall k \in \mathbb{N}. \tag{3.41}$$

Furthermore, since $\tilde{\zeta} > 0$, by (3.39) we also have that $\tilde{\eta}_i > 0 \; \forall i \in I_{\pm}(\hat{y})$. Now since $\left\{ \frac{\eta_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \right\} \longrightarrow \tilde{\eta}_i$, we can then assume w.l.o.g. that $\eta_i^k > 0 \; \forall k \in \mathbb{N}$. This then implies that

$$\eta_i^k = \alpha_k \left(y_i^k - 1 \right) + \bar{\eta}_i^k \quad \forall k \in \mathbb{N}.$$

We then have

$$\begin{aligned} 0 < \tilde{\eta}_i &= \lim_{k \to \infty} \frac{\eta_i^k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} \\ &= \lim_{k \to \infty} \frac{\alpha_k \left(y_i^k - 1 \right)}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} + \lim_{k \to \infty} \frac{\bar{\eta}_i^k}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|} \\ &= \lim_{k \to \infty} \frac{\alpha_k \left(y_i^k - 1 \right)}{\left\| \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k \right) \right\|}, \end{aligned}$$

since $\{\bar{\eta}_i^k\}$ is bounded by definition. Hence, we can assume w.l.o.g. that

$$\alpha_k (y_i^k - 1) > 0 \quad \forall k \in \mathbb{N}$$

and therefore, since $\alpha_k > 0$ we have

$$y_i^k > 1 \quad \forall k \in \mathbb{N}. \tag{3.42}$$

Now by the feasibility of $(\hat{x}, \hat{y})$ we have $\hat{y}_i \leq 1 \quad \forall i \in \{1, \dots, n\}$. From (3.42) we then obtain

$$\hat{y}_i < y_i^k \quad \forall i \in I_{\pm}(\hat{y}) \forall k \in \mathbb{N}. \tag{3.43}$$

Now we claim that

$$\forall k \in \mathbb{N} \exists j_k \in I_0(\hat{y}) : \; y_{j_k}^k \leq 0. \tag{3.44}$$

Suppose not. Then

$$\exists l \in \mathbb{N} \forall i \in I_0(\hat{y}) : \; 0 < y_i^l \quad \iff \quad \exists l \in \mathbb{N} \forall i \in I_0(\hat{y}) : \; \hat{y}_i < y_i^l.$$

From (3.43) we then obtain for l that

$$\hat{y}_i < y_i^l \quad \forall i \in \{1, \dots, n\}$$

and hence,

$$e^T \hat{y} < e^T y^l$$

which contradicts (3.41). Thus, (3.44) holds. Since $\{j_k\} \subseteq I_0(\hat{y})$ and $I_0(\hat{y})$ is a finite set, then there exists $j \in I_0(\hat{y})$ such that $j = j_k$ infinitely often. By passing to a subsequence, we can therefore assume w.l.o.g. that $j = j_k \; \forall k \in \mathbb{N}$. Now since $j \in I_0(\hat{y})$, by (3.35) we have $\eta_j^k = 0 \; \forall k \in \mathbb{N}$ and hence,

$$B_j^k = -\zeta_k + \gamma_j^k x_j^k \quad \iff \quad B_j^k + \zeta_k = \gamma_j^k x_j^k.$$

Since $y_j^k \leq 0$ we then have

$$\gamma_j^k x_j^k = \left(\alpha_k x_j^k y_j^k + \bar{\gamma}_j^k\right) x_j^k = \alpha_k (x_j^k)^2 y_j^k + \bar{\gamma}_j^k x_j^k \leq \bar{\gamma}_j^k x_j^k.$$

Consequently we have

$$B_j^k + \zeta_k \leq \bar{\gamma}_j^k x_j^k$$

and therefore,

$$\frac{B_j^k + \zeta_k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \leq \frac{\bar{\gamma}_j^k x_j^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|}.$$

Since $\{\bar{\gamma}_j^k x_j^k\}$ is bounded, letting $k \longrightarrow \infty$ then yields

$$0 < \zeta \leq 0,$$

which leads to a contradiction. Hence we have

$$(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0.$$

Dividing A^k by $\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|$ and letting $k \longrightarrow \infty$ then yields

$$0 = \sum_{i=1}^{m} \tilde{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \tilde{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \tilde{\gamma}_i e_i$$

where $(\tilde{\lambda}, \tilde{\mu}, \tilde{\gamma}) \neq 0$ and by (3.29), (3.30), (3.34) $\tilde{\lambda} \in \mathbb{R}_+^m$, $\forall i \notin I_g(\hat{x}) : \tilde{\lambda}_i = 0$, as well as $\forall i \in I_{\pm}(\hat{x}) : \tilde{\gamma}_i = 0$.

Now suppose that $i \in \{1, \dots, m\}$ such that $\tilde{\lambda}_i > 0$. Since $\left\{ \frac{\lambda_i^k}{\|(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k)\|} \right\} \longrightarrow \tilde{\lambda}_i$, we can assume w.l.o.g. that $\lambda_i^k > 0 \; \forall k \in \mathbb{N}$ and thus, $\lambda_i^k = \alpha_k g_i(x^k) + \bar{\lambda}_i^k$. Consequently we have

$$0 < \tilde{\lambda}_i = \lim_{k \to \infty} \frac{\lambda_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|}$$

$$
\begin{aligned}
&= \lim_{k\to\infty} \frac{\alpha_k g_i(x^k)}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} + \lim_{k\to\infty} \frac{\bar{\lambda}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\alpha_k g_i(x^k)}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|}
\end{aligned}
$$

by the boundedness of $\{\bar{\lambda}_i^k\}$. Hence, we can assume w.l.o.g. that

$$\alpha_k g_i(x^k) > 0 \quad \forall k \in \mathbb{N}$$

and therefore, since $\alpha_k > 0$

$$g_i(x^k) > 0 \quad \forall k \in \mathbb{N}.$$

This then implies that

$$\tilde{\lambda}_i g_i(x^k) > 0 \quad \forall k \in \mathbb{N}.$$

Now let $i \in \{1, \dots, p\}$ such that $\tilde{\mu}_i \neq 0$. Assume w.l.o.g. that $\tilde{\mu}_i > 0$. The other case can be handled analogously. Here we have by the boundedness of $\{\bar{\mu}_i^k\}$ that

$$
\begin{aligned}
0 < \tilde{\mu}_i &= \lim_{k\to\infty} \frac{\mu_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\alpha_k h_i(x^k)}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} + \lim_{k\to\infty} \frac{\bar{\mu}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\alpha_k h_i(x^k)}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|}.
\end{aligned}
$$

Hence, we can assume w.l.o.g. that

$$\alpha_k h_i(x^k) > 0 \quad \forall k \in \mathbb{N}$$

and therefore, since $\alpha_k > 0$

$$h_i(x^k) > 0 \quad \forall k \in \mathbb{N}.$$

This then implies that

$$\tilde{\mu}_i h_i(x^k) > 0 \quad \forall k \in \mathbb{N}.$$

Suppose now that $i \in \{1, \dots, n\}$ such that $\tilde{\gamma}_i \neq 0$. Assume w.l.o.g. that $\tilde{\gamma}_i > 0$. The other case can be handled analogously. By the boundedness of $\{\bar{\gamma}_i^k\}$ that

$$
\begin{aligned}
0 < \tilde{\gamma}_i &= \lim_{k\to\infty} \frac{\tilde{\gamma}_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\gamma_i^k y_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{(\alpha_k x_i^k y_i^k + \bar{\gamma}_i^k) y_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\alpha_k x_i^k (y_i^k)^2}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} + \lim_{k\to\infty} \frac{\bar{\gamma}_i^k y_i^k}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|} \\
&= \lim_{k\to\infty} \frac{\alpha_k x_i^k (y_i^k)^2}{\left\|\left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right)\right\|}.
\end{aligned}
$$

Hence, we can assume w.l.o.g. that

$$\alpha_k x_i^k (y_i^k)^2 > 0 \quad \forall k \in \mathbb{N}.$$

Consequently

$$x_i^k > 0 \quad \forall k \in \mathbb{N}.$$

and therefore

$$\tilde{\gamma}_i x_i^k > 0 \quad \forall k \in \mathbb{N}.$$

This contradicts the CC-quasinormality of $\hat{x}$. Thus, $\left\{ \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\}$ is bounded and therefore has a convergent subsequence. Assume w.l.o.g. that the whole sequence converges, i.e.

$$\exists \left(\hat{\lambda}, \hat{\mu}, \hat{\gamma}, \hat{\zeta}, \hat{\eta}\right) \,:\, \left\{ \left(\lambda^k, \mu^k, \tilde{\gamma}^k, \zeta^k, \eta^k\right) \right\} \longrightarrow \left(\hat{\lambda}, \hat{\mu}, \hat{\gamma}, \hat{\zeta}, \hat{\eta}\right).$$

Observe that since $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, then we have $\hat{\lambda} \in \mathbb{R}^m_+$. Suppose now that $i \notin I_g(\hat{x})$. Then just like for $\tilde{\lambda}_i$ we can show that $\hat{\lambda}_i = 0$. Similarly, for $i \in I_\pm(\hat{x})$, just like for $\tilde{\gamma}_i$, we can show that $\hat{\gamma}_i = 0$. From the definition of A^k we then obtain by letting $k \longrightarrow 0$ that

$$0 = \nabla f(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \hat{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^n \hat{\gamma}_i e_i$$

where $\forall i \notin I_g(\hat{x}) \,:\, \hat{\lambda}_i = 0$ and $\forall i \in I_\pm(\hat{x}) \,:\, \hat{\gamma}_i = 0$. □

A direct consequence of Proposition 3.56 is the following

Corollary 3.66. *Under the assumptions of Theorem 3.65, there exists $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is an S-stationary point.*

3.4.2 Regularisation Method of Kanzow-Schwartz

We adapt here the regularisation method from [40] for (1.9). Define

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

Let us now collect some important properties of φ. For proofs we refer to [40, Lemma 3.1]

Lemma 3.67 ([40, Lemma 3.1]). *(a) φ is an NCP-function, i.e. $\varphi((a,b)) = 0 \iff a \geq 0, b \geq 0, ab = 0$.*

(b) φ is continuously differentiable with

$$\nabla\varphi((a,b)) = \begin{cases} (b,a)^T & \text{if } a+b \geq 0, \\ (-a,-b)^T & \text{if } a+b < 0. \end{cases}$$

Now let $t > 0$ be a regularisation parameter. In order to relax the constraint $x \circ y = 0$ in (1.9) we define for each $i \in \{1, \dots, n\}$ the following four functions

- $\Phi^{KS}_{1,i}((x,y);t) \,:=\, \varphi((x_i - t, y_i - t)) = \begin{cases} (x_i - t)(y_i - t) & \text{if } x_i + y_i \geq 2t, \\ -\frac{1}{2}((x_i - t)^2 + (y_i - t)^2) & \text{if } x_i + y_i < 2t, \end{cases}$
- $\Phi^{KS}_{2,i}((x,y);t) \,:=\, \varphi((x_i - t, -y_i - t)) = \begin{cases} (x_i - t)(-y_i - t) & \text{if } x_i - y_i \geq 2t, \\ -\frac{1}{2}((x_i - t)^2 + (-y_i - t)^2) & \text{if } x_i - y_i < 2t, \end{cases}$

- $\Phi^{KS}_{3,i}((x,y);t) := \varphi((-x_i - t, -y_i - t)) = \begin{cases} (-x_i - t)(-y_i - t) & \text{if } -x_i - y_i \geq 2t, \\ -\frac{1}{2}((-x_i - t)^2 + (-y_i - t)^2) & \text{if } -x_i - y_i < 2t, \end{cases}$

- $\Phi^{KS}_{4,i}((x,y);t) := \varphi((-x_i - t, y_i - t)) = \begin{cases} (-x_i - t)(y_i - t) & \text{if } -x_i + y_i \geq 2t, \\ -\frac{1}{2}((-x_i - t)^2 + (y_i - t)^2) & \text{if } -x_i + y_i < 2t. \end{cases}$

These functions are continuously differentiable and their derivatives with respect to (x, y) are given in the following lemma.

Lemma 3.68.
- $\nabla\Phi^{KS}_{1,i}((x,y);t) = \begin{cases} \begin{bmatrix} (y_i - t)e_i \\ (x_i - t)e_i \end{bmatrix} & \text{if } x_i + y_i \geq 2t, \\ \begin{bmatrix} (-x_i + t)e_i \\ (-y_i + t)e_i \end{bmatrix} & \text{if } x_i + y_i < 2t, \end{cases}$

- $\nabla\Phi^{KS}_{2,i}((x,y);t) = \begin{cases} \begin{bmatrix} (-y_i - t)e_i \\ -(x_i - t)e_i \end{bmatrix} & \text{if } x_i - y_i \geq 2t, \\ \begin{bmatrix} (-x_i + t)e_i \\ -(y_i + t)e_i \end{bmatrix} & \text{if } x_i - y_i < 2t, \end{cases}$

- $\nabla\Phi^{KS}_{3,i}((x,y);t) = \begin{cases} \begin{bmatrix} (y_i + t)e_i \\ (x_i + t)e_i \end{bmatrix} & \text{if } -x_i - y_i \geq 2t, \\ \begin{bmatrix} -(x_i + t)e_i \\ -(y_i + t)e_i \end{bmatrix} & \text{if } -x_i - y_i < 2t, \end{cases}$

- $\nabla\Phi^{KS}_{4,i}((x,y);t) = \begin{cases} \begin{bmatrix} -(y_i - t)e_i \\ (-x_i - t)e_i \end{bmatrix} & \text{if } -x_i + y_i \geq 2t, \\ \begin{bmatrix} -(x_i + t)e_i \\ (-y_i + t)e_i \end{bmatrix} & \text{if } -x_i + y_i < 2t. \end{cases}$

The proof of the preceeding lemma follows from straightforward computation and is therefore omitted. For $t > 0$ we can now formulate the regularised problem $NLP^{KS}(t)$ as

$$\begin{aligned} \min_{x,y} \ f(x) \quad \text{s.t.} \quad & g_i(x) \leq 0 && \forall i = 1, \dots, m, \\ & h_i(x) = 0 && \forall i = 1, \dots, p, \\ & n - e^T y \leq s, && \\ & y_i \leq 1 && \forall i = 1, \dots, n, \\ & \Phi^{KS}_{j,i}((x,y);t) \leq 0 && \forall i = 1, \dots, n \ \forall j = 1, \dots, 4. \end{aligned} \tag{3.45}$$

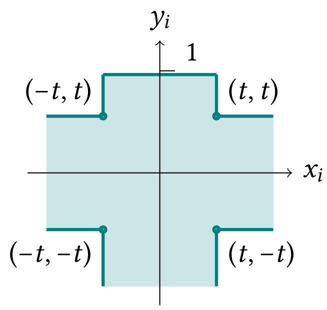


Figure 3.1: Illustration of the Kanzow-Schwartz's relaxation method

Note that our regularised problem slightly differs from the one used in [21] since we drop the constraint $y \geq 0$ here. In the exact case we obtain the following convergence result.

Theorem 3.69. *Let* $\{t_k\} \downarrow 0$, $\hat{x} \in \mathbb{R}^n$, *and* $\{((x^k, y^k), \lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\}$ *be a corresponding sequence of KKT-points of* $NLP^{KS}(t_k)$ *such that* $\{x^k\} \longrightarrow \hat{x}$. *Then* $\hat{x}$ *is a CC-AM-stationary point of* (1.1).

The proof of the preceeding theorem is similar to the inexact case which we will handle next. Hence, we omit it and refer the readers to the proof of Theorem 3.71. Now in order to tackle the inexact case, we first need to define inexactness. Consider (2.1). The following definition of inexactness is taken from [41, Definition 1].

Definition 3.70. *Let* $x \in \mathbb{R}^n$ *and* $\epsilon > 0$. *We then say that* x *is an* ϵ*-stationary point of* (2.1) *iff there exists* $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ *such that*

- $\left\|\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^p \mu_i \nabla h_i(x)\right\| \leq \epsilon$,
- $\lambda_i \geq -\epsilon,\ g_i(x) \leq \epsilon,\ |\lambda_i g_i(x)| \leq \epsilon\ \forall i = 1, \dots, m$,
- $|h_i(x)| \leq \epsilon\ \forall i = 1, \dots, p$.

In the context of MPCC it is known that inexactness destroys the convergence theory of the method, see [41]. This is not the case here.

Theorem 3.71. *Let* $\{t_k\} \downarrow 0$, $\{\epsilon_k\} \downarrow 0$, *and* $\{(x^k, y^k)\}$ *be a sequence of* ϵ_k*-stationary points of* $NLP^{KS}(t_k)$. *Suppose that* $\{x^k\} \longrightarrow \hat{x}$. *Then* $\hat{x}$ *is a CC-AM-stationary point.*

Proof. By assumption, there exists $\{(\lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times (\mathbb{R}^n)^5$ such that

$$(ks_1)\ \left\|\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k)\right\| \leq \epsilon_k,$$

$$(ks_2)\ \left\|-\zeta_k e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_y \Phi_{j,i}^{KS}((x^k, y^k); t_k)\right\| \leq \epsilon_k,$$

$$(ks_3)\ \lambda_i^k \geq -\epsilon_k, \quad g_i(x^k) \leq \epsilon_k, \quad |\lambda_i^k g_i(x^k)| \leq \epsilon_k \quad \forall i = 1, \dots, m,$$

$$(ks_4)\ |h_i(x^k)| \leq \epsilon_k \quad \forall i = 1, \dots, p,$$

$$(ks_5)\ \zeta_k \geq -\epsilon_k, \quad n - e^T y^k - s \leq \epsilon_k, \quad |\zeta_k(n - e^T y^k - s)| \leq \epsilon_k,$$

$$(ks_6)\ \eta_i^k \geq -\epsilon_k, \quad y_i^k - 1 \leq \epsilon_k, \quad |\eta_i^k (y_i^k - 1)| \leq \epsilon_k \quad \forall i = 1, \dots, n,$$

$$(ks_7)\ \gamma_i^{j,k} \geq -\epsilon_k, \quad \Phi_{j,i}^{KS}((x^k, y^k); t_k) \leq \epsilon_k, \quad |\gamma_i^{j,k} \Phi_{j,i}^{KS}((x^k, y^k); t_k)| \leq \epsilon_k \quad \forall j = 1, \dots, 4,\ \forall i = 1, \dots, n.$$

Let us now prove that $\{y^k\}$ is bounded. By (ks_6) we have for each $i \in \{1, \dots, n\}$ that

$$y_i^k \leq 1 + \epsilon_k\ \forall k \in \mathbb{N}.$$

Since $\{1 + \epsilon_k\}$ is a convergent sequence, it is bounded. In particular, there exists $d \in \mathbb{R}$ such that

$$1 + \epsilon_k \leq d\ \forall k \in \mathbb{N}.$$

Thus, it follows that

$$y_i^k \leq d\ \forall k \in \mathbb{N}.$$

We now claim that

$$\forall i \in \{1, \dots, n\} \exists c_i \in \mathbb{R} \forall k \in \mathbb{N}: \ c_i \leq y_i^k. \tag{3.46}$$

Suppose not. Then

$$\exists \hat{i} \in \{1, \dots, n\} \forall c \in \mathbb{R} \exists k \in \mathbb{N} : \ y_{\hat{i}}^k < c.$$

We can thus construct a subsequence $\left\{ y_{\hat{i}}^{k_l} \right\}$ such that $\left\{ y_{\hat{i}}^{k_l} \right\} \longrightarrow -\infty$. On the other hand, by (ks_5) we have

$$n - s - \epsilon_{k_l} \leq e^T y^{k_l} = \sum_{i=1}^{n} y_i^{k_l} = \sum_{i=1, i \neq \hat{i}}^{n} y_i^{k_l} + y_{\hat{i}}^{k_l} \leq (n-1)d + y_{\hat{i}}^{k_l}$$

which leads to a contradiction, since the leftmost part tends to $n - s$ whereas the rightmost part converges to $-\infty$. Thus, (3.46) holds. Define

$$c := \min_{i=1,\dots,n} c_i \ \in \mathbb{R}.$$

Then we have

$$c \leq y_i^k \ \forall i = 1, \dots, n \ \forall k \in \mathbb{N}$$

and therefore

$$\{y^k\} \subseteq [c, d]^n$$

which proves the boundedness of $\{y^k\}$. Since $\{y^k\}$ is bounded, it has a convergent subsequence. By passing to a subsequence we can assume w.l.o.g. that the whole sequence converges, i.e.

$$\exists \hat{y} \in \mathbb{R} : \ \{y^k\} \longrightarrow \hat{y}.$$

In particular, we then have $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$. Let us now prove that $(\hat{x}, \hat{y})$ is feasible for (1.9). By (ks_3) - (ks_6) we obviously have

- $g_i(\hat{x}) \leq 0 \ \forall i = 1, \dots, m,$
- $h_i(\hat{x}) = 0 \ \forall i = 1, \dots, p,$
- $n - e^T \hat{y} \leq s,$
- $\hat{y} \leq e.$

Hence, it remains to prove that $\hat{x} \circ \hat{y} = 0$. Suppose that this is not the case. Then

$$\exists i \in \{1, \dots, n\} : \ \hat{x}_i \hat{y}_i \neq 0.$$

So we need to consider 4 separate cases.
<u>Case 1:</u> $\hat{x}_i > 0 \ \wedge \ \hat{y}_i > 0$
Since $\{x_i^k + y_i^k\} \longrightarrow \hat{x}_i + \hat{y}_i > 0$ and $\{t_k\} \downarrow 0$, we can assume w.l.o.g. that $x_i^k + y_i^k \geq 2t_k \ \forall k \in \mathbb{N}$. Hence we have $\Phi_{1,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(y_i^k - t_k)$. By (ks_7) we then have $\hat{x}_i \hat{y}_i \leq 0$ for the limit which yields a contradiction since $\hat{x}_i \hat{y}_i > 0$ in this case.
<u>Case 2:</u> $\hat{x}_i < 0 \ \wedge \ \hat{y}_i < 0$
Since $\{-x_i^k - y_i^k\} \longrightarrow -\hat{x}_i - \hat{y}_i > 0$ and $\{t_k\} \downarrow 0$, we can assume w.l.o.g. that $-x_i^k - y_i^k \geq 2t_k \ \forall k \in \mathbb{N}$. Hence we have $\Phi_{3,i}^{KS}((x^k, y^k); t_k) = (-x_i^k - t_k)(-y_i^k - t_k)$. By (ks_7) we then have $\hat{x}_i \hat{y}_i \leq 0$ for the limit which yields a contradiction since $\hat{x}_i \hat{y}_i > 0$ in this case.
<u>Case 3:</u> $\hat{x}_i > 0 \ \wedge \ \hat{y}_i < 0$
Since $\{x_i^k - y_i^k\} \longrightarrow \hat{x}_i - \hat{y}_i > 0$ and $\{t_k\} \downarrow 0$, we can assume w.l.o.g. that $x_i^k - y_i^k \geq 2t_k \ \forall k \in \mathbb{N}$. Hence we have $\Phi_{2,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(-y_i^k - t_k)$. By (ks_7) we then have $-\hat{x}_i \hat{y}_i \leq 0$ for the limit which yields a contradiction since $-\hat{x}_i \hat{y}_i > 0$ in this case.
<u>Case 4:</u> $\hat{x}_i < 0 \ \wedge \ \hat{y}_i > 0$
Since $\{-x_i^k + y_i^k\} \longrightarrow -\hat{x}_i + \hat{y}_i > 0$ and $\{t_k\} \downarrow 0$, we can assume w.l.o.g. that $-x_i^k + y_i^k \geq 2t_k \ \forall k \in \mathbb{N}$. Hence we have $\Phi_{4,i}^{KS}((x^k, y^k); t_k) = (-x_i^k - t_k)(y_i^k - t_k)$. By (ks_7) we then have $-\hat{x}_i \hat{y}_i \leq 0$ for the limit which yields a contradiction since $-\hat{x}_i \hat{y}_i > 0$ in this case.

Hence we can conclude that $\hat{x}\circ\hat{y} = 0$ and therefore, $(\hat{x}, \hat{y})$ is feasible for (1.9). By Theorem 3.44 $\hat{x}$ is then feasible for (1.1). Now define

$$w^k := \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k). \tag{3.47}$$

By (ks_1) we know that $\{w^k\} \longrightarrow 0$. Suppose now that $i \notin I_g(\hat{x})$. Since $\{g_i(x^k)\} \longrightarrow g_i(\hat{x}) < 0$, we can assume w.l.o.g. that

$$g_i(x^k) < 0 \;\forall k \in \mathbb{N}.$$

This implies that

$$|g_i(x^k)| > 0 \;\forall k \in \mathbb{N}$$

and hence,

$$0 \leq |\lambda_i^k| \leq \frac{\epsilon_k}{|g_i(x^k)|} \;\forall k \in \mathbb{N}.$$

Letting $k \longrightarrow \infty$ we then obtain $\{\lambda_i^k\} \longrightarrow 0$ and therefore, $\{\lambda_i^k \nabla g_i(x^k)\} \longrightarrow 0$. Reformulating (3.47) we then obtain for each $k \in \mathbb{N}$

$$\begin{aligned} w^k - \sum_{i\notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) = \nabla f(x^k) &+ \sum_{i\in I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \end{aligned} \tag{3.48}$$

where the left hand side tends to 0. Now define for each $k \in \mathbb{N}$

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k + \epsilon_k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{else.} \end{cases} \tag{3.49}$$

By (ks_3) we then have $\{\hat{\lambda}^k\} \subseteq \mathbb{R}_+^m$. Since $\{\epsilon_k\} \longrightarrow 0$ we then have $\{\epsilon_k \nabla g_i(x^k)\} \longrightarrow 0$ for each $i \in I_g(\hat{x})$ as well. Reformulating (3.48) then yields

$$\begin{aligned} w^k - \sum_{i\notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i\in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k) &= \nabla f(x^k) + \sum_{i\in I_g(\hat{x})} (\lambda_i^k + \epsilon_k) \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \end{aligned}$$

where the left hand side converges to 0. Suppose now that $i \in I_\pm(\hat{x})$. By the feasibility of $(\hat{x}, \hat{y})$ for (1.9) we then have $\hat{y}_i = 0$. Assume first that $\hat{x}_i > 0$. Since $\{t_k\} \downarrow 0$ and $\{\text{sgn}1 x_i^k + \text{sgn}2 y_i^k\} \longrightarrow \text{sgn}1\hat{x}_i + \text{sgn}2\hat{y}_i = \text{sgn}1\hat{x}_i$ where sgn1, sgn2 $\in \{+, -\}$, we can assume w.l.o.g. that for each $k \in \mathbb{N}$ we have

- $x_i^k + y_i^k \geq 2t_k$
  $\Longrightarrow \Phi_{1,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(y_i^k - t_k)$
  $\Longrightarrow \nabla_x \Phi_{1,i}^{KS}((x^k, y^k); t_k) = (y_i^k - t_k)e_i$,

- $x_i^k - y_i^k \geq 2t_k$
  $\Longrightarrow \Phi_{2,i}^{KS}((x^k, y^k); t_k) = (x_i^k - t_k)(-y_i^k - t_k)$
  $\Longrightarrow \nabla_x \Phi_{2,i}^{KS}((x^k, y^k); t_k) = (-y_i^k - t_k)e_i$,
- $-x_i^k - y_i^k < 2t_k$
  $\Longrightarrow \Phi_{3,i}^{KS}((x^k, y^k); t_k) = -\frac{1}{2}((-x_i^k - t_k)^2 + (-y_i^k - t_k)^2)$
  $\Longrightarrow \nabla_x \Phi_{3,i}^{KS}((x^k, y^k); t_k) = -(x_i^k + t_k)e_i$,
- $-x_i^k + y_i^k < 2t_k$
  $\Longrightarrow \Phi_{4,i}^{KS}((x^k, y^k); t_k) = -\frac{1}{2}((-x_i^k - t_k)^2 + (y_i^k - t_k)^2)$
  $\Longrightarrow \nabla_x \Phi_{4,i}^{KS}((x^k, y^k); t_k) = -(x_i^k + t_k)e_i$.

By (ks_7) we have

- $|\gamma_i^{1,k} \Phi_{1,i}^{KS}((x^k, y^k); t_k)| = |\gamma_i^{1,k}(x_i^k - t_k)(y_i^k - t_k)| \leq \epsilon_k$.
  Now since $\{x_i^k - t_k\} \longrightarrow \hat{x}_i > 0$, we can assume w.l.o.g. that $|x_i^k - t_k| > 0 \; \forall k \in \mathbb{N}$. We then have

  $$0 \leq |\gamma_i^{1,k}(y_i^k - t_k)| \leq \frac{\epsilon_k}{|x_i^k - t_k|}.$$

  By sandwich theorem we obtain

  $$\{\gamma_i^{1,k}(y_i^k - t_k)\} \longrightarrow 0 \quad \Longrightarrow \quad \{\gamma_i^{1,k} \nabla_x \Phi_{1,i}^{KS}((x^k, y^k); t_k)\} \longrightarrow 0.$$

- $|\gamma_i^{2,k} \Phi_{2,i}^{KS}((x^k, y^k); t_k)| = |\gamma_i^{2,k}(x_i^k - t_k)(-y_i^k - t_k)| \leq \epsilon_k$.
  Using a similar argument as before we obtain

  $$\{\gamma_i^{2,k}(-y_i^k - t_k)\} \longrightarrow 0 \quad \Longrightarrow \quad \{\gamma_i^{2,k} \nabla_x \Phi_{2,i}^{KS}((x^k, y^k); t_k)\} \longrightarrow 0.$$

- $|\gamma_i^{3,k} \Phi_{3,i}^{KS}((x^k, y^k); t_k)| = |\gamma_i^{3,k}(-\frac{1}{2})((-x_i^k - t_k)^2 + (-y_i^k - t_k)^2)| \leq \epsilon_k$.
  To simplify the notation let

  $$\beta_k := | -\frac{1}{2}((-x_i^k - t_k)^2 + (-y_i^k - t_k)^2)|.$$

  Observe that $\{\beta_k\} \longrightarrow \frac{1}{2}\hat{x}_i^2 > 0$. Hence we can assume w.l.o.g. that $\beta_k > 0$ for each $k \in \mathbb{N}$. We then have

  $$0 \leq |\gamma_i^{3,k}| \leq \frac{\epsilon_k}{\beta_k}.$$

  Applying sandwich theorem yields

  $$\{\gamma_i^{3,k}\} \longrightarrow 0 \quad \Longrightarrow \quad \{\gamma_i^{3,k} \nabla_x \Phi_{3,i}^{KS}((x^k, y^k); t_k)\} = \{\gamma_i^{3,k}(-(x_i^k + t_k))e_i\} \longrightarrow 0.$$

- $|\gamma_i^{4,k} \Phi_{4,i}^{KS}((x^k, y^k); t_k)| = |\gamma_i^{4,k}(-\frac{1}{2})((-x_i^k - t_k)^2 + (y_i^k - t_k)^2)| \leq \epsilon_k$.
  Using a similar argument as before we obtain that

  $$\{\gamma_i^{4,k}\} \longrightarrow 0 \quad \Longrightarrow \quad \{\gamma_i^{4,k} \nabla_x \Phi_{4,i}^{KS}((x^k, y^k); t_k)\} \longrightarrow 0.$$

Similarly, for the case where $\hat{x}_i < 0$ we can also prove that

$$\{\gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k)\} \longrightarrow 0 \; \forall j = 1, \ldots, 4.$$

Putting things together we obtain

$$\lim_{k \to \infty} \sum_{i \in I_{\pm}(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) = 0.$$

Defining

$$A^k := w^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k) - \sum_{i \in I_\pm(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k)$$

for each $k \in \mathbb{N}$ we obtain

$$A^k = \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k)$$

and $\{A^k\} \longrightarrow 0$. Now by Lemma 3.68 we know that for each $i \in I_0(\hat{x})$ we have

$$\nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \in \operatorname{span}\{e_i\} \quad \forall j = 1, \ldots, 4$$

and hence,

$$\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) \in \operatorname{span}\{e_i\}.$$

In particular, there exists then $\hat{\gamma}_i^k \in \mathbb{R}$ such that

$$\sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{KS}((x^k, y^k); t_k) = \hat{\gamma}_i^k e_i.$$

Then we have

$$A^k = \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \hat{\gamma}_i^k e_i.$$

Now define for each $i \in I_\pm(\hat{x})$ $\hat{\gamma}_i^k := 0$. Then we have for each $k \in \mathbb{N}$

$$A^k = \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \hat{\gamma}_i^k e_i.$$

By (3.49) and since $\{A^k\} \longrightarrow 0$, it then follows that $\hat{x}$ is a CC-AM-stationary point. □

An immediate consequence of Theorem 3.71 and Theorem 3.35 is the following

Corollary 3.72. *If, in addition to the assumptions in Theorem 3.71, $\hat{x}$ also satisfies CC-AM-regularity, then $\hat{x}$ is CC-M-stationary.*

Proposition 3.56 then immediately implies the following

Corollary 3.73. *Under the assumptions of Corollary 3.72, there exists $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

In light of Corollary 3.42, the result obtained here is stronger than the result from [21], not only because we take inexactness into account, but also because in [21] the authors used a stronger constraint qualification.

3.4.3 Regularisation Method of Steffensen-Ulbrich

Let us now adapt the regularisation method from [54] for (1.9). In the context of MPCC it is known that in the exact case this method theoretically has a weaker convergence property compared to the regularisation method of Kanzow-Schwartz, see [36]. As we shall see later, this is not the case here. The method from [54] is theoretically just as strong as the method from [40]. We begin with a definition.

Definition 3.74 ([54, Assumption 3.1]). *Let $D \subseteq \mathbb{R}$ be an open set containing $[-1,1]$. A function $\theta : D \longrightarrow \mathbb{R}$ is called a regularisation function iff it satisfies the following properties:*

(a) θ *is twice continuously differentiable on* $[-1,1]$,

(b) $\theta(-1) = \theta(1) = 1$,

(c) $\theta'(-1) = -1,\ \theta'(1) = 1$,

(d) $\theta''(-1) = \theta''(1) = 0$,

(e) θ *is strictly convex on* $[-1,1]$.

Two examples of such functions are

$$\theta(z) := \frac{2}{\pi}\sin\left(\frac{\pi}{2}z + \frac{3\pi}{2}\right) + 1 \quad \text{and} \quad \theta(z) := \frac{1}{8}(-z^4 + 6z^2 + 3),$$

where the second function is the Hermite interpolation polynomial satisfying the requirements from Definition 3.74, see also [35]. We recall now a useful property of a regularisation function.

Lemma 3.75 ([54, Lemma 3.1]). *Let θ be a regularisation function. Then we have for each $z \in (-1,1)$ that $\theta(z) > |z|$.*

Suppose now that we have a regularisation function θ. Let $t > 0$ be a regularisation parameter. Following [54] we next define

$$\varphi(\ ;t) : \mathbb{R} \longrightarrow \mathbb{R},\ \varphi(u;t) := \begin{cases} |u| & \text{if } |u| \geq t, \\ t\theta\left(\frac{u}{t}\right) & \text{if } |u| < t. \end{cases}$$

Let us gather some properties of this function.

Lemma 3.76 ([35, Lemma 4.4]). *The following properties hold for φ*

(a) $\varphi(u;t) > |u|\ \forall u \in (-t,t)\ \forall t > 0$,

(b) $\varphi(u;t) = |u|\ \forall |u| \geq t\ \forall t > 0$,

(c) $\lim_{t \longrightarrow 0} \varphi(u;t) = |u|\ \forall u \in \mathbb{R}$,

(d) $\varphi(\ ;t)$ *is twice continuously differentiable for all* $t > 0$.

In order to relax the orthogonality constraint $x \circ y = 0$, we define for each $i \in \{1,\dots,n\}$ the following four functions

- $\Phi^{SU}_{1,i}((x,y);t) := x_i + y_i - \varphi(x_i - y_i;t)$,
- $\Phi^{SU}_{2,i}((x,y);t) := x_i - y_i - \varphi(x_i + y_i;t)$,
- $\Phi^{SU}_{3,i}((x,y);t) := -x_i - y_i - \varphi(-x_i + y_i;t)$,
- $\Phi^{SU}_{4,i}((x,y);t) := -x_i + y_i - \varphi(-x_i - y_i;t)$.

Straightforward computations then yield

Lemma 3.77.

- $$\Phi^{SU}_{1,i}((x,y);t) = \begin{cases} 2y_i & \textit{if } x_i - y_i \geq t, \\ 2x_i & \textit{if } x_i - y_i \leq -t, \\ x_i + y_i - t\theta\left(\frac{x_i - y_i}{t}\right) & \textit{if } x_i - y_i \in (-t,t), \end{cases}$$

$$\nabla\Phi^{SU}_{1,i}((x,y);t) = \begin{cases} \begin{bmatrix} 0 \\ 2e_i \end{bmatrix} & \textit{if } x_i - y_i \geq t, \\ \begin{bmatrix} 2e_i \\ 0 \end{bmatrix} & \textit{if } x_i - y_i \leq -t, \\ \begin{bmatrix} (1 - \theta'\left(\frac{x_i - y_i}{t}\right))e_i \\ (1 + \theta'\left(\frac{x_i - y_i}{t}\right))e_i \end{bmatrix} & \textit{if } x_i - y_i \in (-t,t), \end{cases}$$

- $$\Phi^{SU}_{2,i}((x,y);t) = \begin{cases} -2y_i & \textit{if } x_i + y_i \geq t, \\ 2x_i & \textit{if } x_i + y_i \leq -t, \\ x_i - y_i - t\theta\left(\frac{x_i + y_i}{t}\right) & \textit{if } x_i + y_i \in (-t,t), \end{cases}$$

$$\nabla\Phi^{SU}_{2,i}((x,y);t) = \begin{cases} \begin{bmatrix} 0 \\ -2e_i \end{bmatrix} & \textit{if } x_i + y_i \geq t, \\ \begin{bmatrix} 2e_i \\ 0 \end{bmatrix} & \textit{if } x_i + y_i \leq -t, \\ \begin{bmatrix} (1 - \theta'\left(\frac{x_i + y_i}{t}\right))e_i \\ (-1 - \theta'\left(\frac{x_i + y_i}{t}\right))e_i \end{bmatrix} & \textit{if } x_i + y_i \in (-t,t), \end{cases}$$

- $$\Phi^{SU}_{3,i}((x,y);t) = \begin{cases} -2y_i & \textit{if } -x_i + y_i \geq t, \\ -2x_i & \textit{if } -x_i + y_i \leq -t, \\ -x_i - y_i - t\theta\left(\frac{-x_i + y_i}{t}\right) & \textit{if } -x_i + y_i \in (-t,t), \end{cases}$$

$$\nabla\Phi^{SU}_{3,i}((x,y);t) = \begin{cases} \begin{bmatrix} 0 \\ -2e_i \end{bmatrix} & \textit{if } -x_i + y_i \geq t, \\ \begin{bmatrix} -2e_i \\ 0 \end{bmatrix} & \textit{if } -x_i + y_i \leq -t, \\ \begin{bmatrix} (-1 + \theta'\left(\frac{-x_i + y_i}{t}\right))e_i \\ (-1 - \theta'\left(\frac{-x_i + y_i}{t}\right))e_i \end{bmatrix} & \textit{if } -x_i + y_i \in (-t,t), \end{cases}$$

- $$\Phi^{SU}_{4,i}((x,y);t) = \begin{cases} 2y_i & \textit{if } -x_i - y_i \geq t, \\ -2x_i & \textit{if } -x_i - y_i \leq -t, \\ -x_i + y_i - t\theta\left(\frac{-x_i - y_i}{t}\right) & \textit{if } -x_i - y_i \in (-t,t), \end{cases}$$

$$\nabla\Phi^{SU}_{4,i}((x,y);t) = \begin{cases} \begin{bmatrix} 0 \\ 2e_i \end{bmatrix} & \textit{if } -x_i - y_i \geq t, \\ \begin{bmatrix} -2e_i \\ 0 \end{bmatrix} & \textit{if } -x_i - y_i \leq -t, \\ \begin{bmatrix} (-1 + \theta'\left(\frac{-x_i - y_i}{t}\right))e_i \\ (1 + \theta'\left(\frac{-x_i - y_i}{t}\right))e_i \end{bmatrix} & \textit{if } -x_i - y_i \in (-t,t). \end{cases}$$

We can now formulate the regularised problem $NLP^{SU}(t)$ as

$$\begin{aligned}\min_{x,y}\ f(x) \quad \text{s.t.} \quad & g_i(x) \leq 0 && \forall i = 1, \dots, m,\\ & h_i(x) = 0 && \forall i = 1, \dots, p,\\ & n - e^T y \leq s, && \\ & y_i \leq 1 && \forall i = 1, \dots, n,\\ & \Phi^{SU}_{j,i}((x, y); t) \leq 0 && \forall i = 1, \dots, n\ \forall j = 1, \dots, 4.\end{aligned} \tag{3.50}$$

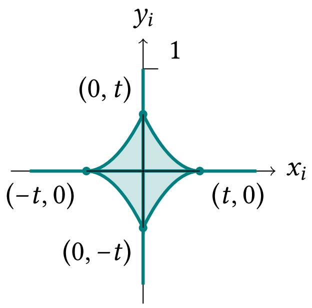


Figure 3.2: Illustration of the Steffensen-Ulbrich's relaxation method

For the exact case we obtain the following convergence result, which parallels the result obtained for the regularisation method of Kanzow-Schwartz.

Theorem 3.78. *Let* $\{t_k\} \downarrow 0$ *and* $\{((x^k, y^k), \lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\}$ *be a corresponding sequence of KKT-points of* $NLP^{SU}(t_k)$ *such that* $\{x^k\} \longrightarrow \hat{x}$*. Then* $\hat{x}$ *is a CC-AM-stationary point of* (1.1).

The proof of the theorem is similar to the inexact case. Hence, we omit it and refer the readers to the proof of Theorem 3.79. As we shall see, unlike in MPCC [41], here the method of Steffensen-Ulbrich retains its convergence property in the inexact case as well.

Theorem 3.79. *Let* $\{t_k\} \downarrow 0$, $\{\epsilon_k\} \downarrow 0$, *and* $\{(x^k, y^k)\}$ *be a sequence of* ϵ_k*-stationary points of* $NLP^{SU}(t_k)$*. Suppose that* $\{x^k\} \longrightarrow \hat{x}$*. Then* $\hat{x}$ *is a CC-AM-stationary point.*

Proof. By assumption, there exists $\{(\lambda^k, \mu^k, \zeta_k, \eta^k, \gamma^{1,k}, \gamma^{2,k}, \gamma^{3,k}, \gamma^{4,k})\} \subseteq \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times (\mathbb{R}^n)^5$ such that

$$(su_1)\quad \left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_x \Phi^{SU}_{j,i}((x^k, y^k); t_k) \right\| \leq \epsilon_k,$$

$$(su_2)\quad \left\| -\zeta_k e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_y \Phi^{SU}_{j,i}((x^k, y^k); t_k) \right\| \leq \epsilon_k,$$

$$(su_3)\quad \lambda_i^k \geq -\epsilon_k, \quad g_i(x^k) \leq \epsilon_k, \quad |\lambda_i^k g_i(x^k)| \leq \epsilon_k \quad \forall i = 1, \dots, m,$$

$$(su_4)\quad |h_i(x^k)| \leq \epsilon_k \quad \forall i = 1, \dots, p,$$

$$(su_5)\quad \zeta_k \geq -\epsilon_k, \quad n - e^T y^k - s \leq \epsilon_k, \quad |\zeta_k(n - e^T y^k - s)| \leq \epsilon_k,$$

$$(su_6)\quad \eta_i^k \geq -\epsilon_k, \quad y_i^k - 1 \leq \epsilon_k, \quad |\eta_i^k (y_i^k - 1)| \leq \epsilon_k \quad \forall i = 1, \dots, n,$$

$$(su_7)\quad \gamma_i^{j,k} \geq -\epsilon_k, \quad \Phi^{SU}_{j,i}((x^k, y^k); t_k) \leq \epsilon_k, \quad |\gamma_i^{j,k} \Phi^{SU}_{j,i}((x^k, y^k); t_k)| \leq \epsilon_k \quad \forall j = 1, \dots, 4,\ \forall i = 1, \dots, n.$$

Using the same argument as in the proof of Theorem 3.71, we can show that $\{y^k\}$ is bounded. Hence, we can assume w.l.o.g. that there exists $\hat{y} \in \mathbb{R}^n$ such that $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$. Let us now prove that $(\hat{x}, \hat{y})$ is feasible for (1.9). By (su_3) - (su_6) we clearly have

$$g(\hat{x}) \leq 0, \quad h(\hat{x}) = 0, \quad n - e^T \hat{y} \leq s, \quad \hat{y} \leq e.$$

Hence, it remains to show that $\hat{x} \circ \hat{y} = 0$. Suppose not. Then

$$\exists i \in \{1, \dots, n\} \ : \ \hat{x}_i \hat{y}_i \neq 0.$$

We thus have four cases to handle, namely

$$x_i > 0 \ \wedge \ y_i > 0, \quad x_i < 0 \ \wedge \ y_i < 0, \quad x_i > 0 \ \wedge \ y_i < 0, \quad x_i < 0 \ \wedge \ y_i > 0.$$

Here we shall derive a contradiction only for the case where $\hat{x}_i > 0 \ \wedge \ \hat{y}_i > 0$. The other three cases can be dealt with analogously. Since $\{\epsilon_k\} \downarrow 0$, $\{x_i^k\} \longrightarrow \hat{x}_i > 0$, and $\{y_i^k\} \longrightarrow \hat{y}_i > 0$ we can assume w.l.o.g. that

$$x_i^k > \frac{\epsilon_k}{2} \quad \wedge \quad y_i^k > \frac{\epsilon_k}{2} \quad \forall k \in \mathbb{N}. \tag{3.51}$$

We now claim that

$$|x_i^k - y_i^k| < t_k \ \forall k \in \mathbb{N}. \tag{3.52}$$

Suppose not. Then

$$\exists l \in \mathbb{N} \ : \ |x_i^l - y_i^l| \geq t_l.$$

This then implies that

$$\varphi(x_i^l - y_i^l; t_l) = |x_i^l - y_i^l|$$

and hence, by (su_7) and (3.51)

$$\epsilon_l \geq \Phi_{1,i}^{SU}((x^l, y^l); t_l) = x_i^l + y_i^l - |x_i^l - y_i^l| = 2\min\{x_i^l, y_i^l\} > \epsilon_l,$$

which is a contradiction. Thus, (3.52) holds. Consequently, by (su_7) and Lemma 3.77, we have for each $k \in \mathbb{N}$

$$\epsilon_k \geq \Phi_{1,i}^{SU}((x^k, y^k); t_k) = x_i^k + y_i^k - t_k \theta\left(\frac{x_i^k - y_i^k}{t_k}.\right) \tag{3.53}$$

Now by (3.52) the sequence $\left\{\frac{x_i^k - y_i^k}{t_k}\right\}$ is clearly bounded and therefore, it has a convergent subsequence. We can thus assume w.l.o.g. that the whole sequence converges, i.e.

$$\exists a \in \mathbb{R} \ : \ \left\{\frac{x_i^k - y_i^k}{t_k}\right\} \longrightarrow a.$$

Then letting $k \longrightarrow \infty$ we obtain from (3.53) by the continuity of θ

$$0 \geq \hat{x}_i + \hat{y}_i - 0 \cdot \theta(a) = \hat{x}_i + \hat{y}_i > 0.$$

This leads to a contradiction. Thus, we conclude that $\hat{x} \circ \hat{y} = 0$ and hence, $(\hat{x}, \hat{y})$ is feasible for (1.9). By Theorem 3.44 $\hat{x}$ is then feasible for (1.1). Now similar to the proof of Theorem 3.71, we can define $\{\hat{\lambda}^k\} \subseteq \mathbb{R}_+^m$ where

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k + \epsilon_k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{else} \end{cases}$$

for each $k \in \mathbb{N}$. Defining

$$w^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \sum_{j=1}^4 \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{SU}((x^k, y^k); t_k)$$

just like in the proof of Theorem 3.71 we then obtain

$$w^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k) = \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k)$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{SU}((x^k, y^k); t_k)$$

where the left hand side converges to 0. Suppose now that $i \in I_{\pm}(\hat{x})$. Since $(\hat{x}, \hat{y})$ is feasible for (1.9) we then have $\hat{y}_i = 0$. Assume first that $\hat{x}_i > 0$. Since $\{x_i^k\} \longrightarrow \hat{x}_i$, we can assume w.l.o.g. that $x_i^k > 0 \; \forall k \in \mathbb{N}$. Furthermore, since $\{t_k\} \downarrow 0$ and $\{\text{sgn1} x_i^k + \text{sgn2} y_i^k\} \longrightarrow \text{sgn1}\hat{x}_i + \text{sgn2}\hat{y}_i = \text{sgn1}\hat{x}_i$ where sgn1, sgn2 $\in \{+, -\}$, by Lemma 3.77 we can assume w.l.o.g. that for each $k \in \mathbb{N}$ we have

- $x_i^k - y_i^k \geq t_k$
  $\Longrightarrow \nabla_x \Phi_{1,i}^{SU}((x^k, y^k); t_k) = 0,$
  $\Longrightarrow \{\gamma_i^{1,k} \nabla_x \Phi_{1,i}^{SU}((x^k, y^k); t_k)\} = \{0\} \longrightarrow 0,$
- $x_i^k + y_i^k \geq t_k$
  $\Longrightarrow \nabla_x \Phi_{2,i}^{SU}((x^k, y^k); t_k) = 0,$
  $\Longrightarrow \{\gamma_i^{2,k} \nabla_x \Phi_{2,i}^{SU}((x^k, y^k); t_k)\} = \{0\} \longrightarrow 0,$
- $-x_i^k + y_i^k \leq -t_k$
  $\Longrightarrow \Phi_{3,i}^{SU}((x^k, y^k); t_k) = -2x_i^k.$
  By (su_7) we then have
  $$|\gamma_i^{3,k}(-2x_i^k)| \leq \epsilon_k \quad \Longrightarrow \quad |\gamma_i^{3,k}| \leq \frac{\epsilon_k}{2x_i^k}.$$
  Letting $k \longrightarrow \infty$ then yields
  $$\{\gamma_i^{3,k}\} \longrightarrow 0 \quad \Longrightarrow \quad \{\gamma_i^{3,k} \nabla_x \Phi_{3,i}^{SU}((x^k, y^k); t_k)\} = \{\gamma_i^{3,k}(-2e_i)\} \longrightarrow 0.$$
- $-x_i^k - y_i^k \leq -t_k$
  Using a similar argument as in the previous case then yields
  $$\{\gamma_i^{4,k} \nabla_x \Phi_{4,i}^{SU}((x^k, y^k); t_k)\} \longrightarrow 0.$$

Similarly we can also prove that for $\hat{x}_i < 0$ we have

$$\{\gamma_i^{j,k} \nabla_x \Phi_{j,i}^{SU}((x^k, y^k); t_k)\} \longrightarrow 0 \; \forall j = 1, \dots, 4.$$

Putting things together we obtain

$$\lim_{k \to \infty} \sum_{i \in I_{\pm}(\hat{x})} \sum_{j=1}^{4} \gamma_i^{j,k} \nabla_x \Phi_{j,i}^{SU}((x^k, y^k); t_k) = 0.$$

The rest of the proof is then essentially the same as in the proof of Theorem 3.71. We can therefore conclude that $\hat{x}$ is a CC-AM-stationary point. □

As a direct consequence of Theorem 3.79 and Theorem 3.35 we obtain the following

Corollary 3.80. *If, in addition to the assumptions in Theorem 3.79, $\hat{x}$ also satisfies CC-AM-regularity, then $\hat{x}$ is CC-M-stationary.*

Proposition 3.56 then implies the following.

Corollary 3.81. *Under the assumptions of Corollary 3.80, there exists $\hat{z} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{z})$ is CC-S-stationary.*

3.5 Numerical Experiments

In this section we shall compare the performance of ALGENCAN with the Scholtes regularisation method from [18] as well as the Kanzow-Schwartz regularisation method from [21]. All experiments were conducted using Python together with the Numpy library. We used ALGENCAN 2.4.0 compiled with MA57 library [37] and called through its Python interface with user-supplied gradients of the objective functions, sparse Jacobian of the constraints, as well as sparse Hessian of the Lagrangian. As a subsolver for the regularisation methods of Scholtes and Kanzow-Schwartz we used the for academic use freely available ESA SQP solver WORHP version 1.14 [22] called through its Python interface. For the Scholtes regularisation method WORHP was called with user-supplied sparse gradients of the objective functions, sparse Jacobian of the constraints, as well as the sparse Hessian of the Lagrangian. On the other hand, for the Kanzow-Schwartz regularisation method, since the analytical Hessian does not exist as the corresponding NCP-function is not twice differentiable, we called WORHP with user-supplied sparse gradients of the objective functions and sparse Jacobian of the constraints only. The Hessian of the Lagrangian was then approximated using the BFGS method. Throughout the experiments both ALGENCAN and WORHP were called using their respective default settings. We applied ALGENCAN directly to the relaxed reformulations of the test problems as in (1.9), i.e. without the lower bound for the auxiliary variable y. In contrast, following [18] and [21], for both regularisation methods we bounded y from below by 0. Next, for each test problem we started both regularisation methods with an initial regularisation parameter $t_0 = 1.0$ and decreased t_k in each iteration by a factor of 0.01. The regularisation methods were terminated if either $t_k < 10^{-8}$ or $\left\| x^k \circ y^k \right\|_\infty \leq 10^{-6}$.

3.5.1 Pilot Test

Let us begin by considering the following academic example

$$\min_{x \in \mathbb{R}^2} x_1 + 10x_2 \quad \text{s.t.} \quad \left(x_1 - \frac{1}{2}\right)^2 + (x_2 - 1)^2 \leq 1, \ \|x\|_0 \leq 1$$

which is taken from [21]. This problem has a local minimiser in $\left(0, 1 - \frac{1}{2}\sqrt{3}\right)$ and an isolated global minimiser in $\left(\frac{1}{2}, 0\right)$. Following [21], we discretised the rectangle $\left[-1, \frac{3}{2}\right] \times \left[-\frac{1}{2}, 2\right]$ resulting in 441 starting points for the considered methods. For each of these starting points ALGENCAN converged towards the global minimiser $\left(\frac{1}{2}, 0\right)$. The same behaviour was also observed for the Scholtes regularisation method. On the other hand, the Kanzow-Schwartz regularisation method was slightly less succesful, converging in 437 cases towards the global minimiser. In the other 4 cases the method converged towards the local minimiser. This behaviour might be due to the performance of the BFGS method used by WORHP in approximating the Hessian of the Lagrangian. Indeed, running the Scholtes regularisation method without user-supplied Hessian of the Lagrangian, letting the Hessian be approximated by the BFGS method instead, yielded in a convergence towards the global minimiser in only 394 cases. In the other 47 cases the Scholtes regularisation method only managed to find the local minimiser.

3.5.2 Portfolio Optimisation Problems

Following [21] we consider a classical portfolio optimisation problem

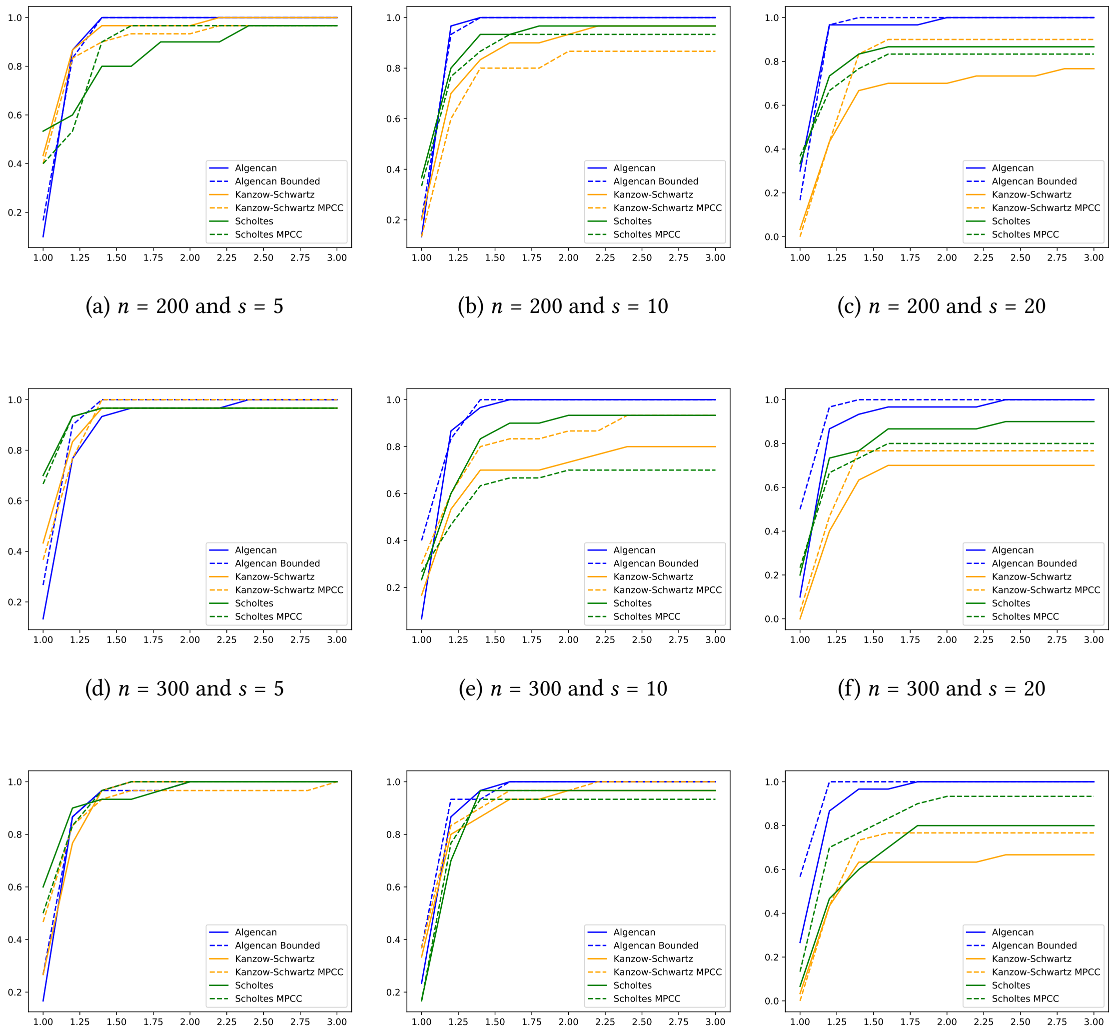
$$\min_{x \in \mathbb{R}^n} x^T Q x \quad \text{s.t.} \quad \mu^T x \geq \rho, \ e^T x \leq 1, \ 0 \leq x \leq u, \ \|x\|_0 \leq s, \tag{3.54}$$

where Q and μ are the covariance matrix as well as the mean of n possible assets and $e^T x \leq 1$ is the budget constraint, see [14, 26]. We generated the test problems using the data from [32], considering $s = 5, 10, 20$ for each dimension $n = 200, 300, 400$ which resulted in 270 test problems, see also [21]. Note that due to the constraint $x \geq 0$ in (3.54), if we include the lower bound 0 for y in the relaxed reformulation as in (A.1), then the feasible set actually has the classical MPCC structure. Thus, for each regularisation method, one regularisation function in the first quadrant actually already suffices. Moreover, for comparison purpose, we also applied ALGENCAN to the relaxed reformulations of the problems as in (A.1). Hence, we considered a total of six approaches:

- ALGENCAN without a lower bound on y
- ALGENCAN with an additional lower bound $y \geq 0$
- Scholtes and Kanzow-Schwartz regularisation for cardinality-constrained problems [18, 21] with a regularisation of both upper quadrants
- Scholtes and Kanzow-Schwartz regularisation for MPCCs [40, 53] with a regularisation of the upper right quadrant only.

For each test problem, we used the initial values $x^0 = 0$ and $y^0 = e$. As a performance measure for the considered methods we compared the attained objective function values and generated a performance profile as suggested in [27], where we set the objective function value of a method for a problem to be ∞ if the method failed to find a feasible point of the problem within a tolerance of 10^{-6}.

As can be seen from Figure 3.3, ALGENCAN worked very reliably with regard to the feasibility of the solutions. It often outperformed the regularisation methods in terms of the objective function value of the solution, especially for larger values of s. Lastly we note that although introducing the lower bound $y \geq 0$ does not have any theoretical effect on ALGENCAN, the numerical results suggest that it could bring slight improvements to ALGENCAN's performance.



(a) $n = 200$ and $s = 5$ (b) $n = 200$ and $s = 10$ (c) $n = 200$ and $s = 20$

(d) $n = 300$ and $s = 5$ (e) $n = 300$ and $s = 10$ (f) $n = 300$ and $s = 20$

(g) $n = 400$ and $s = 5$ (h) $n = 400$ and $s = 10$ (i) $n = 400$ and $s = 20$

Figure 3.3: Comparing the performance of ALGENCAN and the regularisation methods for (3.54)

Chapter 4

Sparse Optimisation Problems

In this chapter we shall deal with (1.2). Some of the results in this chapter, particularly those concerning the equivalence between the minima of (1.2) and (1.10), are taken by the author from the unpublished manuscript [20] by Oleg P. Burdakov, Christian Kanzow, and Alexandra Schwartz. It is one of the aims of this thesis to extend [20] by proving some additional results, in particular those pertaining to sequential optimality conditions and the numerical behaviour of the considered algorithms. The results which are taken from the original manuscript [20] have been labelled as such in this thesis and the author would like to take this opportunity to express his gratitude to Oleg P. Burdakov, Christian Kanzow, and Alexandra Schwartz for coming up with these important results and permitting the author to use them in his work.

Let us begin by investigating how (1.2) relates to (1.1). In our subsequent analyses we would also let s to be in $\{0, n\}$.

Theorem 4.1. *Let $\hat{x} \in \mathbb{R}^n$ be a local minimiser of* (1.1). *Then it is also a local minimiser of* (1.2).

Proof. By assumption $\hat{x} \in X \cap \mathcal{S}$ and there exists $\epsilon_1 > 0$ such that for each $x \in X \cap \mathcal{S} \cap B_{\epsilon_1}(\hat{x})$ we have $f(\hat{x}) \leq f(x)$. Furthermore, by the lower semicontinuity of f there exists $\epsilon_2 > 0$ such that for each $x \in B_{\epsilon_2}(\hat{x})$ we have $f(\hat{x}) < f(x) + \rho$. Moreover, by Lemma 3.6, there exists $\epsilon_3 > 0$ such that for each $x \in B_{\epsilon_3}(\hat{x})$ we have $\|\hat{x}\|_0 \leq \|x\|_0$. Define $\epsilon := \min_{i=1,2,3} \epsilon_i > 0$. Now let $x \in X \cap B_\epsilon(\hat{x})$. Then $x \in B_{\epsilon_3}(\hat{x})$. We differentiate between 2 cases.

Case 1: $\|x\|_0 = \|\hat{x}\|_0$

Since by assumption $\|\hat{x}\|_0 \leq s$, we then also have $x \in \mathcal{S}$. Thus, $x \in X \cap \mathcal{S} \cap B_{\epsilon_1}(\hat{x})$. This implies that $f(\hat{x}) \leq f(x)$ and hence,

$$f(\hat{x}) + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|\hat{x}\|_0 = f(x) + \rho\|x\|_0.$$

Case 2: $\|x\|_0 > \|\hat{x}\|_0$

By the definition of $\|\cdot\|_0$ we then have $\|\hat{x}\|_0 + 1 \leq \|x\|_0$. Since $x \in B_{\epsilon_2}(\hat{x})$ we then have

$$f(\hat{x}) + \rho\|\hat{x}\|_0 < f(x) + \rho + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|x\|_0. \qquad \square$$

Observe that Theorem 4.1 holds irrespective of the choice of ρ and s. However, it should be noted that a global minimiser of (1.1) is not necessarily a global minimiser of (1.2).

Example 4.2 ([20, Example 3.1]). *Consider the cardinality-constrained problem*

$$\min \ f(x) := \|Ax - b\|^2 \quad s.t. \quad \|x\|_0 \leq 2$$

where

$$A := \begin{bmatrix} 0 & 3 & -3 \\ 3 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \quad \wedge \quad b := \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}.$$

The optimal value of f respective to the value of $\|\cdot\|_0$ is given as follows:

$\|\cdot\|_0$	0	1	2	3
Opt. value of f	36	18	≈ 11.08	0

Due to the cardinality constraint, the optimal value of the problem is then ≈ 11.08 *and for the corresponding global minimiser* $\hat{x}$ *we have* $\|\hat{x}\|_0 = 2$. *Now for a given sparsity parameter* $\rho > 0$, *consider the sparse optimisation problem*

$$\min \; F_\rho(x) := f(x) + \rho\|x\|_0.$$

In light of the above table, the optimal value of F_ρ *respective to the value of* $\|\cdot\|_0$ *is then given by*

$\|\cdot\|_0$	0	1	2	3
Opt. value of F_ρ	36	$18+\rho$	$\approx 11.08 + 2\rho$	3ρ

Suppose that $\rho < \approx 11.08$. *Then clearly we have* $3\rho < \approx 11.08 + 2\rho$ *and hence, since* $\|\hat{x}\|_0 = 2$, $\hat{x}$ *cannot be a global minimiser of* F_ρ. *Suppose now that* $\rho \geq \approx 11.08$. *Then*

$$\approx 11.08 + 2\rho - (18+\rho) = \rho - \approx 6.92 > 0.$$

Hence, $\hat{x}$ *cannot be a global minimiser of* F_ρ. *Thus, we conclude that there is no* $\rho > 0$ *such that* $\hat{x}$ *is a global minimiser of* F_ρ.

Suppose that $0 \in X$. Letting $s := 0$ in (1.1), we then clearly have for each $x \in \mathbb{R}^n$ that $\|x\|_0 \leq s$ iff $x = 0$. Hence, 0 is the only feasible point of (1.1) and thus, a global minimiser of (1.1). Inspecting the proof of Theorem 4.1 we then immediately obtain the following result.

Proposition 4.3. *Let* $0 \in X$. *Then it is a strict local minimiser of* (1.2).

Now let us consider (2.1). Letting $s := n$ in (1.1), obviously (2.1) is then equivalent to (1.1). Theorem 4.1 subsequently implies the following.

Proposition 4.4. *Let* $\hat{x} \in X$ *be a local minimiser of* (2.1). *Then* $\hat{x}$ *is also a local minimiser of* (1.2).

Note, however, that a global minimiser of (2.1) is not necessarily a global minimiser of (1.2).

Example 4.5. *Consider*

$$\min_{x\in\mathbb{R}} \left(x - \frac{1}{2}\right)^2. \tag{4.1}$$

Obviously $\frac{1}{2}$ *is the only global minimiser of this problem. However, it is not a global minimiser of*

$$\min_{x\in\mathbb{R}} \left(x - \frac{1}{2}\right)^2 + \|x\|_0. \tag{4.2}$$

Indeed, it is easy to see that 0 *is the only global minimiser of* (4.2).

Moreover, the converse of Proposition 4.4 is false in general.

Example 4.6. *Consider again* (4.1) *and* (4.2). *As we have seen before,* 0 *is a global minimiser of* (4.2). *However, it is obviously not a stationary point of* (4.1). *Thus, it cannot be a local minimiser of* (4.1).

As a direct consequence of Proposition 4.4 we obtain the following

Corollary 4.7. *Let* $\hat{x} \in X$. *Then* $\hat{x}$ *is a local minimiser of*

$$\min_{x\in\mathbb{R}^n} \rho\|x\|_0 \quad \text{s.t.} \quad x \in X. \tag{4.3}$$

Proof. We consider the special case where $f \equiv 0$, the constant zero function. Then every feasible point $\hat{x} \in X$ is a local minimiser of (2.1). Therefore, by Proposition 4.4, it is also a local minimiser of (4.3). □

Example 4.8. *Corollary 4.7 immediately implies that every feasible point of the compressive sensing problem*

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \quad s.t. \quad Ax \geq b, \quad Cx = d,$$

where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{k\times n}$, $b \in \mathbb{R}^m$, *and* $d \in \mathbb{R}^k$ *for* $k, m, n \in \mathbb{N}$ *is a local minimiser of the problem.*

This extends the result obtained in [30, page 283].

For the converse of Theorem 4.1 we obtain the following two results.

Proposition 4.9. *Let* $\hat{x} \in X$ *be a local minimiser of* (1.2). *Then for* $s := \|\hat{x}\|_0$ $\hat{x}$ *is also a local minimiser of* (1.1).

Proof. By assumption, there exists $\epsilon_1 > 0$ such that for each $x \in X \cap B_{\epsilon_1}(\hat{x})$ we have $f(\hat{x}) + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|x\|_0$. Moreover, by Lemma 3.6 there exists $\epsilon_2 > 0$ such that for each $x \in B_{\epsilon_2}(\hat{x})$ we have $\|\hat{x}\|_0 \leq \|x\|_0$. Define $\epsilon := \min_{i=1,2} \epsilon_i > 0$. Then for each $x \in X \cap \mathcal{S} \cap B_\epsilon(\hat{x})$ we have

$$\|\hat{x}\|_0 \overset{\epsilon\leq\epsilon_2}{\leq} \|x\|_0 \leq s = \|\hat{x}\|_0 \Longrightarrow \|x\|_0 = \|\hat{x}\|_0 \Longrightarrow f(\hat{x}) + \rho\|\hat{x}\|_0 \overset{\epsilon\leq\epsilon_1}{\leq} f(x) + \rho\|x\|_0 = f(x) + \rho\|\hat{x}\|_0 \Longrightarrow f(\hat{x}) \leq f(x). \quad \square$$

Proposition 4.10 ([20, Proposition 3.2])**.** *Let* $\hat{x} \in X$ *be a global minimiser of* (1.2). *Then for* $s := \|\hat{x}\|_0$ $\hat{x}$ *is also a global minimiser of* (1.1).

Proof. By assumption we have for each $x \in X$ that $f(\hat{x}) + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|x\|_0$. Now let $x \in X \cap \mathcal{S}$. Then we have

$$f(\hat{x}) + \rho\|\hat{x}\|_0 \overset{x\in X}{\leq} f(x) + \rho\|x\|_0 \overset{\|x\|_0\leq s}{\leq} f(x) + \rho s \overset{s=\|\hat{x}\|_0}{=} f(x) + \rho\|\hat{x}\|_0 \Longrightarrow f(\hat{x}) \leq f(x). \quad \square$$

Note that if $s = n$, then (1.1) is equivalent to (2.1). Proposition 4.9 and Proposition 4.10 then lead to the following result.

Proposition 4.11. *Let* $\hat{x} \in X$ *be a local (global) minimiser of* (1.2) *such that* $\|\hat{x}\|_0 = n$. *Then it is a local (global) minimiser of* (2.1).

4.1 Sequential Optimality Conditions

The mapping $\rho\|\cdot\|_0$ is obviously lower semicontinuous. Let us now compute its Fréchet subdifferential.

Definition 4.12 ([29, Theorem 5.2.11])**.** *Let* $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ *and* $\hat{x} \in \mathbb{R}^n$. *Then the Fréchet subdifferential of* ϕ *at* $\hat{x}$ *is defined as*

$$\partial^F \phi(\hat{x}) := \left\{ \gamma \in \mathbb{R}^n \,\middle|\, \liminf_{h\longrightarrow 0, h\neq 0} \frac{\phi(\hat{x}+h) - \phi(\hat{x}) - \gamma^T h}{\|h\|} \geq 0 \right\}.$$

Lemma 4.13. *Let* $\hat{x} \in \mathbb{R}^n$. *Then*

$$\partial^F(\rho\|\hat{x}\|_0) = \{\gamma \in \mathbb{R}^n \mid \gamma_i = 0 \;\forall i \in I_\pm(\hat{x})\}.$$

Proof. "$\subseteq$": Suppose that $\gamma \in \partial^F(\rho\|\hat{x}\|_0)$. Let $i \in I_\pm(\hat{x})$. Pick a sequence $\{\delta_k\} \subseteq \mathbb{R}_+$ such that $\{\delta_k\} \downarrow 0$. Since $\hat{x}_i \neq 0$ and $\{\delta_k\} \downarrow 0$, we can assume w.l.o.g. that for each $k \in \mathbb{N}$ we have $|\hat{x}_i| > \frac{\delta_k}{2}$. By the reverse triangle inequality we then have

$$\left|\hat{x}_i + sgn\frac{\delta_k}{2}\right| \geq |\hat{x}_i| - \left|sgn\frac{\delta_k}{2}\right| = |\hat{x}_i| - \frac{\delta_k}{2} > 0 \quad \forall k \in \mathbb{N}$$

where $sgn \in \{+,-\}$. In particular, this implies that

$$\hat{x}_i + sgn\frac{\delta_k}{2} \neq 0 \quad \Longleftrightarrow \quad i \in I_\pm\left(\hat{x} + sgn\frac{\delta_k}{2}e_i\right) \quad \Longrightarrow \quad \left\|\hat{x} + sgn\frac{\delta_k}{2}e_i\right\|_0 = \|\hat{x}\|_0 \tag{4.4}$$

for each $k \in \mathbb{N}$. Now observe that for each $k \in \mathbb{N}$ we also have $sgn\frac{\delta_k}{2}e_i \in B_{\delta_k}(0)$. Hence,

$$\begin{aligned}\inf_{h \in B_{\delta_k}(0)\setminus\{0\}} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} &\leq \frac{\rho\left\|\hat{x} + \frac{\delta_k}{2}e_i\right\|_0 - \rho\|\hat{x}\|_0 - sgn\frac{\delta_k}{2}\gamma^T e_i}{\left\|sgn\frac{\delta_k}{2}e_i\right\|} \\ &\overset{(4.4)}{=} \frac{-sgn\frac{\delta_k}{2}\gamma_i}{\frac{\delta_k}{2}} \\ &= -sgn\gamma_i. \end{aligned} \tag{4.5}$$

Letting $k \longrightarrow \infty$ we then obtain

$$\begin{aligned} 0 &\leq \liminf_{h\longrightarrow 0, h\neq 0} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \\ &\overset{[52,\text{Definition 1.5}]}{=} \lim_{k\longrightarrow\infty}\left(\inf_{h \in B_{\delta_k}(0)\setminus\{0\}} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|}\right) \\ &\overset{(4.5)}{\leq} -sgn\gamma_i. \end{aligned}$$

Since $sgn \in \{+,-\}$, this implies that

$$\gamma_i \geq 0 \ \wedge \ -\gamma_i \geq 0 \quad \Longleftrightarrow \quad \gamma_i = 0.$$

"$\supseteq$": Let $\gamma \in \mathbb{R}^n$ such that $\gamma_i = 0 \ \forall i \in I_\pm(\hat{x})$. Suppose that $\epsilon > 0$ be given as in Lemma 3.6. Now pick a sequence $\{\delta_k\} \subseteq \mathbb{R}_+$ such that $\{\delta_k\} \downarrow 0$. Since $\epsilon > 0$ and $\{\delta_k\} \downarrow 0$, we can assume w.l.o.g. that $\delta_k \leq \epsilon$ for each $k \in \mathbb{N}$. We differentiate between two cases.

Case 1: $\gamma = 0$

Let $k \in \mathbb{N}$ and $h \in B_{\delta_k}(0) \setminus \{0\}$. Clearly we have $\hat{x} + h \in B_{\delta_k}(\hat{x}) \subseteq B_\epsilon(\hat{x})$ and hence, by Lemma 3.6, $\rho\|\hat{x}+h\|_0 \geq \rho\|\hat{x}\|_0$. This implies that

$$\frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \overset{\gamma=0}{\geq} 0.$$

Since this holds for each $h \in B_{\delta_k}(0) \setminus \{0\}$, we then have

$$\inf_{h \in B_{\delta_k}(0)\setminus\{0\}} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \geq 0.$$

Letting $k \longrightarrow \infty$ then yields

$$\begin{aligned}\liminf_{h\longrightarrow 0, h\neq 0} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} &\overset{[52,\text{Definition 1.5}]}{=} \lim_{k\longrightarrow\infty}\left(\inf_{h \in B_{\delta_k}(0)\setminus\{0\}} \frac{\rho\|\hat{x}+h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|}\right) \\ &\geq 0. \end{aligned}$$

Case 2: $\gamma \neq 0$

In this case we have $\|\gamma\| > 0$. Since $\{\delta_k\} \downarrow 0$ we can then assume w.l.o.g. that $\delta_k < \frac{\rho}{\|\gamma\|}$ for each $k \in \mathbb{N}$. Now let $k \in \mathbb{N}$ and $h \in B_{\delta_k}(0) \setminus \{0\}$. By Cauchy-Schwarz inequality we obtain

$$\left|\gamma^T h\right| \leq \|\gamma\|\|h\| < \|\gamma\|\delta_k < \rho \quad \Longrightarrow \quad \gamma^T h < \rho \tag{4.6}$$

Since $\delta_k \leq \epsilon$, by Lemma 3.6 we have

$$I_\pm(\hat{x}) \subseteq I_\pm(\hat{x} + h).$$

Case 2.1: $I_\pm(\hat{x}) = I_\pm(\hat{x} + h)$
Here we have

$$\|\hat{x}\|_0 = \|\hat{x} + h\|_0 \quad \wedge \quad I_0(\hat{x}) = I_0(\hat{x} + h).$$

Now let $i \in I_0(\hat{x})$, Then, since $I_0(\hat{x}) = I_0(\hat{x} + h)$,

$$0 \overset{i \in I_0(\hat{x}+h)}{=} (\hat{x} + h)_i = \hat{x}_i + h_i \overset{i \in I_0(\hat{x})}{=} h_i. \tag{4.7}$$

Thus,

$$\gamma^T h = \sum_{i=1}^n \gamma_i h_i = \sum_{i \in I_\pm(\hat{x})} \gamma_i h_i + \sum_{i \in I_0(\hat{x})} \gamma_i h_i = 0$$

by the definition of γ and (4.7). This then implies that

$$\frac{\rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} = 0.$$

Case 2.2: $I_\pm(\hat{x}) \subsetneq I_\pm(\hat{x} + h)$
Since $I_\pm(\hat{x} + h)$ contains at least one more element compared to $I_\pm(\hat{x})$ we then have

$$\|\hat{x} + h\|_0 \geq \|\hat{x}\|_0 + 1 \quad \Longrightarrow \quad \rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 \geq \rho.$$

This then implies that

$$\frac{\rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \geq \frac{\rho - \gamma^T h}{\|h\|} \overset{(4.6)}{>} 0.$$

In both cases we have

$$\frac{\rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \geq 0$$

and therefore

$$\inf_{h \in B_{\delta_k}(0) \setminus \{0\}} \frac{\rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \geq 0.$$

Letting $k \longrightarrow \infty$ we then obtain by [52, Definition 1.5]

$$\liminf_{h \longrightarrow 0, h \neq 0} \frac{\rho\|\hat{x} + h\|_0 - \rho\|\hat{x}\|_0 - \gamma^T h}{\|h\|} \geq 0.$$

Thus, we conclude that $\gamma \in \partial^F(\rho\|\hat{x}\|_0))$. □

Motivated by the CAKKT-condition for (2.1) and the CC-CAM-stationarity for (1.1) we introduce the following definition.

Definition 4.14. *Let $\hat{x} \in X$. We say that $\hat{x}$ is an* SP complementary approximately Karush-Kuhn-Tucker (SP-CAKKT) point *iff there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} \longrightarrow 0,$

(c) $\left\{\sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\gamma_i^k x_i^k|\right\} \longrightarrow 0.$

The next theorem states that SP-CAKKT is a first-order necessary optimality condition for (1.2).

Theorem 4.15. *Let $\hat{x} \in X$ be a local minimiser of* (1.2). *Then $\hat{x}$ is an SP-CAKKT point.*

Proof. By assumption, there exists $\epsilon > 0$ such that

$$f(\hat{x}) + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|x\|_0 \quad \forall x \in \bar{B}_\epsilon(\hat{x}) \cap X.$$

In particular, $\hat{x}$ is then the unique global minimiser of

$$\min_x \; f(x) + \rho\|x\|_0 + \frac{1}{2}\|x - \hat{x}\|^2 \quad \text{s.t.} \quad x \in \bar{B}_\epsilon(\hat{x}) \cap X. \tag{4.8}$$

Now pick a sequence $\{\alpha_k\} \subseteq \mathbb{R}_+$ such that $\{\alpha_k\} \uparrow \infty$ and consider for each $k \in \mathbb{N}$ the following penalised problem

$$\min_x \; f(x) + \frac{\alpha_k}{2}\|(g(x)_+, h(x)\|^2 + \frac{1}{2}\|x - \hat{x}\|^2 + \rho\|x\|_0 \quad \text{s.t.} \quad x \in \bar{B}_\epsilon(\hat{x}). \tag{4.9}$$

Observe that $x \in X$ iff $\|(g(x)_+, h(x)\| = 0$. Moreover, the objective function of (4.9) is lower semicontinuous for each $k \in \mathbb{N}$. Furthermore, the feasible set $\bar{B}_\epsilon(\hat{x})$ is compact. Hence, by [29, Theorem 2.5.3], for each $k \in \mathbb{N}$ (4.9) admits a global minimiser $x^k \in \bar{B}_\epsilon(\hat{x})$. Now since $\hat{x} \in \bar{B}_\epsilon(\hat{x}) \cap X$ we then have for each $k \in \mathbb{N}$

$$\begin{aligned} f(x^k) + \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k)\|^2 + \frac{1}{2}\left\|x^k - \hat{x}\right\|^2 &\overset{\rho\|x^k\|_0 \geq 0}{\leq} f(x^k) + \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k)\|^2 + \frac{1}{2}\left\|x^k - \hat{x}\right\|^2 \\ &\qquad + \rho\|x^k\|_0 \\ &\leq f(\hat{x}) + \rho\|\hat{x}\|_0. \end{aligned} \tag{4.10}$$

Moreover, by the compactness of $\bar{B}_\epsilon(\hat{x})$ the sequence $\{x^k\}$ has a convergent subsequence in $\bar{B}_\epsilon(\hat{x})$. Assume w.l.o.g. that the whole sequence converges, i.e. $\exists \bar{x} \in \bar{B}_\epsilon(\hat{x}) : \{x^k\} \longrightarrow \bar{x}$. We shall now show that $\bar{x} = \hat{x}$. Dividing both sides of (4.10) by α_k and taking the limit as $k \longrightarrow \infty$ yields $0 \leq \|(g(\bar{x})_+, h(\bar{x}))\| \leq 0$. This implies that $\bar{x} \in X \cap \bar{B}_\epsilon(\hat{x})$ and therefore, it is feasible for (4.8). Furthermore, we also obtain from (4.10) that

$$f(x^k) + \frac{1}{2}\left\|x^k - \hat{x}\right\|^2 + \rho\|x^k\|_0 \leq f(\hat{x}) + \rho\|\hat{x}\|_0$$

and hence,

$$\begin{aligned} f(\bar{x}) + \frac{1}{2}\|\bar{x} - \hat{x}\|^2 + \rho\|\bar{x}\|_0 &\overset{[29,\text{Theorem 2.5.2}]}{\leq} f(\bar{x}) + \frac{1}{2}\|\bar{x} - \hat{x}\|^2 + \liminf_{k\longrightarrow\infty} \rho\|x^k\|_0 \\ &= \lim_{k\longrightarrow\infty}\left(f(x^k) + \frac{1}{2}\|x^k - \hat{x}\|^2\right) + \liminf_{k\longrightarrow\infty} \rho\|x^k\|_0 \\ &\overset{[2,\text{Aufgabe 2d Abschnitt II.5}]}{=} \liminf_{k\longrightarrow\infty} f(x^k) + \frac{1}{2}\|x^k - \hat{x}\|^2 + \rho\|x^k\|_0 \\ &\overset{[2,\text{Aufgabe 2e Abschnitt II.5}]}{\leq} \liminf_{k\longrightarrow\infty} f(\hat{x}) + \rho\|\hat{x}\|_0 \\ &= f(\hat{x}) + \rho\|\hat{x}\|_0 + \frac{1}{2}\|\hat{x} - \hat{x}\|_0. \end{aligned}$$

But since $\hat{x}$ is the unique global minimiser of (4.8), it follows that $\bar{x} = \hat{x}$. Hence we have $\{x^k\} \longrightarrow \hat{x}$. We can then assume w.l.o.g. that $x^k \in B_\epsilon(\hat{x})$ for each $k \in \mathbb{N}$. Then for each $k \in \mathbb{N}$ x^k is a local minimiser of

$$\min_x \; f(x) + \frac{\alpha_k}{2}\|(g(x)_+, h(x))\|^2 + \frac{1}{2}\|x - \hat{x}\|^2 + \rho\|x\|_0.$$

Hence,

$$0 \overset{[29,\text{Theorem 5.2.23}]}{\in} \partial^F\left(\left(f(x^k) + \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2\right) + \frac{1}{2}\|x^k - \hat{x}\|^2\right) + \rho\|x^k\|_0\Big)$$

$$\overset{[29,\text{Proposition 5.2.30}]}{=} \nabla\left(f(x^k) + \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2) + \frac{1}{2}\|x^k - \hat{x}\|^2\right) + \partial^F(\rho\|x^k\|_0)$$

$$= \nabla f(x^k) + \sum_{i=1}^m \left(\alpha_k \max\{0, g_i(x^k)\}\right)\nabla g_i(x^k) + \sum_{i=1}^p \left(\alpha_k h_i(x^k)\right)\nabla h_i(x^k) + x^k - \hat{x} + \partial^F(\rho\|x^k\|_0). \quad (4.11)$$

Define for each $k \in \mathbb{N}$

$$\begin{aligned} \lambda_i^k &:= \alpha_k \max\{0, g_i(x^k)\} \quad \forall i = 1,\dots,m,\\ \mu_i^k &:= \alpha_k h_i(x^k) \quad \forall i = 1,\dots,p. \end{aligned}$$

Then (4.11) can be rewritten as

$$0 \in \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k) + x^k - \hat{x} + \partial^F(\rho\|x^k\|_0).$$

Thus, for each $k \in \mathbb{N}$ there exists $\gamma^k \in \partial^F(\rho\|x^k\|_0)$ such that

$$0 = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k) + x^k - \hat{x} + \gamma^k$$

which is equivalent to

$$\hat{x} - x^k = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i.$$

Since $\{x^k\} \longrightarrow \hat{x}$ we then have

$$\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k g_i(x^k) + \sum_{i=1}^p \mu_i^k h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} \longrightarrow 0.$$

Observe that by definition we have $\{\lambda^k\} \subseteq \mathbb{R}^m_+$. Now let $k \in \mathbb{N}$. By Lemma 4.13 we then have for each $i \in I_\pm(x^k)$ that $\gamma_i^k = 0$ and hence, $|\gamma_i^k x_i^k| = 0$. Moreover, we also have for each $i \in I_0(x^k)$ that $x_i^k = 0$ and therefore, $|\gamma_i^k x_i^k| = 0$. This then implies that for each $k \in \mathbb{N}$ we have

$$\sum_{i=1}^n |\gamma_i^k x_i^k| = 0.$$

Now (4.10) also implies that

$$f(x^k) + \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 + \rho\|x^k\|_0 \le f(\hat{x}) + \rho\|\hat{x}\|_0,$$

which is equivalent to

$$\frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 + \rho\|x^k\|_0 \le f(\hat{x}) - f(x^k) + \rho\|\hat{x}\|_0.$$

This then implies that

$$\begin{aligned} \liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 + \rho\|\hat{x}\|_0 &\overset{[29,\text{Theorem 2.5.2}]}{\le} \liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 + \liminf_{k\to\infty} \rho\|x^k\|_0\\ &\overset{[2,\text{Aufgabe 2b Abschnitt II.5}]}{\le} \liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 + \rho\|x^k\|_0\\ &\overset{[2,\text{Aufgabe 2e Abschnitt II.5}]}{\le} \liminf_{k\to\infty} f(\hat{x}) - f(x^k) + \rho\|\hat{x}\|_0\\ &\overset{[2,\text{Theorem 5.7}]}{=} \rho\|\hat{x}\|_0 \end{aligned}$$

and hence,

$$\liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 \le 0.$$

On the other hand, since $0 \le \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 \ \forall k \in \mathbb{N}$, by [2, Aufgabe 2e Abschnitt II.5] we also have that

$$0 \le \liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2.$$

Thus,

$$\liminf_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 = 0.$$

By [2, Theorem 5.5, Satz 1.17] there exists a subsequence of $\left\{ \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 \right\}$ which converges to 0. Hence, by passing to this subsequence we can assume w.l.o.g. that

$$\lim_{k\to\infty} \frac{\alpha_k}{2}\|(g(x^k)_+, h(x^k))\|^2 = 0 \implies \lim_{k\to\infty} \alpha_k\|(g(x^k)_+, h(x^k))\|^2 = 0.$$

The rest of the proof is then essentially the same as in the proof of Theorem 3.9 and we can conclude that $\hat{x}$ is an SP-CAKKT point. □

Remark 4.16. *Note that for each $s \ge \|\hat{x}\|_0$ $\hat{x} \in X$ is an SP-CAKKT point of* (1.2) *iff it is a CC-CAM-stationary point of* (1.1).

In light of Theorem 4.1 and Remark 4.16, by fixing an arbitrary $\rho > 0$ we can view Theorem 3.9 as a corollary of Theorem 4.15. On the other hand, the proofs of [21, Theorem 3.4] and Theorem 3.9 do not actually prohibit s to be in $\{0, n\}$. Thus, in light of Proposition 4.9 and Remark 4.16, by setting $s := \|\hat{x}\|_0$, Theorem 4.15 can be viewed as a corollary of Theorem 3.9.

The converse of Theorem 4.15 is false in general as the next example shows.

Example 4.17. *Consider the following problem which is adapted from [8]*

$$\min_{x\in\mathbb{R}^2} \ 3(x_1 - 1) - 2(x_2 - 1) + \|x\|_0 \quad s.t. \quad (x_1 - 1) - (x_2 - 1)\exp(x_2 - 1) \le 0, \ x_1 - x_2 = 0. \tag{4.12}$$

$\hat{x} := (1,1)^T$ is obviously feasible for (4.12). *Let us now show that it is an SP-CAKKT point. Define for each $k \in \mathbb{N}$ $x^k := (1 + 1/k, 1 + 1/k)^T$, $\lambda_k := (\exp(1/k) + (1/k)\exp(1/k) - 1)^{-1}$, $\mu_k := -3 - \lambda_k$, $\gamma^k := 0$. Clearly we have $\{x^k\} \longrightarrow \hat{x}$. Moreover, since $1/k > 0$ for each $k \in \mathbb{N}$, by the strict monotonicity of* exp *we have* $\exp(1/k) > 1$ *and hence, in particular,* $\exp(1/k) + (1/k)\exp(1/k) - 1 > 0$. *Thus, $\{\lambda_k\} \subseteq \mathbb{R}_+$. Now we also have for each $k \in \mathbb{N}$*

$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} + \lambda_k \begin{bmatrix} 1 \\ -\exp(1/k) - (1/k)\exp(1/k) \end{bmatrix} + \mu_k \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \gamma^k = \begin{bmatrix} 3 \\ -2 \end{bmatrix} + \lambda_k \begin{bmatrix} 1 \\ -\lambda_k^{-1} - 1 \end{bmatrix} + \mu_k \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.$$

Furthermore, $|\mu_k(x_1^k - x_2^k)| = 0 \ \forall k \in \mathbb{N}$. Note that for each $k \in \mathbb{N}$ since $\exp(1/k) > 1$ *we also have*

$$\begin{aligned} |\lambda_k(1/k - (1/k)\exp(1/k))| &= \left| \frac{1/k - (1/k)\exp(1/k)}{\exp(1/k) + (1/k)\exp(1/k) - 1} \right| \\ &= \left| \frac{1 - \exp(1/k)}{k\exp(1/k) + \exp(1/k) - k} \right| \\ &= \frac{\exp(1/k) - 1}{k(\exp(1/k) - 1) + \exp(1/k)} \\ &\le \frac{\exp(1/k) - 1}{k(\exp(1/k) - 1)} \\ &= \frac{1}{k}. \end{aligned}$$

Thus, it follows that $\{|\lambda_k(1/k - (1/k)\exp(1/k))|\} \longrightarrow 0$. *Hence, we conclude that* $\hat{x}$ *is an SP-CAKKT point. We now claim that* $\hat{x}$ *is not a local minimiser of* (4.12). *Suppose to the contrary that it is a local minimiser of* (4.12). *Then, since* $\|\hat{x}\|_0 = 2$, *by Proposition 4.11* $\hat{x}$ *must be a local minimiser of the following problem*

$$\min_{x\in\mathbb{R}^2} 3(x_1 - 1) - 2(x_2 - 1) \quad s.t. \quad (x_1 - 1) - (x_2 - 1)\exp(x_2 - 1) \le 0,\ x_1 - x_2 = 0. \tag{4.13}$$

The objective function value of $\hat{x}$ *for* (4.13) *is 0. On the other hand, for each* $k \in \mathbb{N}$ $(1 - 1/k, 1 - 1/k)^T$ *is also feasible for* (4.13) *since*

$$-1/k < 0 \implies \exp(-1/k) < 1 \implies (1/k)\exp(-1/k) < 1/k \implies -1/k + (1/k)\exp(-1/k) < 0.$$

Moreover, the objective function value of $(1-1/k, 1-1k)^T$ *for* (4.13) *is* $-1/k < 0$. *Since* $\{(1-1/k, 1-1/k)^T\} \longrightarrow (1,1)^T$, *this contradicts the assumption that* $\hat{x}$ *is a local minimiser of* (4.13). *Thus, we conclude that* $\hat{x}$ *is not a local minimiser of* (4.12).

Nevertheless, under some additional assumptions we can prove the following result.

Theorem 4.18. *Assume that in* (1.2) *the functions* f *as well as* $g_1, \dots, g_m$ *are convex and* $h_1, \dots, h_p$ *are affine-linear. Let* $\hat{x} \in X$. *If it is an SP-CAKKT point, then it is also a local minimiser of* (1.2).

Proof. Note that the proof of Theorem 3.11 also holds for $s \in \{0, n\}$. By defining $s := \|\hat{x}\|_0$, $\hat{x}$ is then feasible for (1.1). Since $\hat{x}$ is an SP-CAKKT point of (1.2), it is then also a CC-CAM-stationary point of (1.1). The assertion then follows from Theorem 3.11 and Theorem 4.1. □

Motivated by the AKKT condition for (2.1) and the CC-AM-stationarity for (1.1) we introduce the following.

Definition 4.19. *Let* $\hat{x} \in X$. *We say that* $\hat{x}$ *is an* SP approximately Karush-Kuhn-Tucker (SP-AKKT) *point iff there exist sequences* $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, *and* $\{\gamma^k\} \subseteq \mathbb{R}^n$ *such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} \longrightarrow 0$,

(c) $\forall i \notin I_g(\hat{x}) :\ \lambda_i^k = 0\ \forall k \in \mathbb{N}$,

(d) $\forall i \in I_\pm(\hat{x}) :\ \gamma_i^k = 0\ \forall k \in \mathbb{N}$.

Remark 4.20. *Note that for each* $s \ge \|\hat{x}\|_0$ $\hat{x} \in X$ *is an SP-AKKT point of* (1.2) *iff it is a CC-AM-stationary point of* (1.1).

Since the proof of Theorem 3.14 also holds for $s \in \{0, n\}$, by defining $s := \|\hat{x}\|_0$ we immediately obtain the following result from Remark 4.16 and Remark 4.20.

Theorem 4.21. *Let* $\hat{x} \in X$ *be an SP-CAKKT point of* (1.2). *Then it is also an SP-AKKT point of* (1.2).

The converse of Theorem 4.21 is false in general as the next example shows.

Example 4.22. *Consider the following problem*

$$\min_{x\in\mathbb{R}^3} \frac{(x_2 - 2)^2}{2} + \|x\|_0 \quad s.t. \quad x_1 x_2 = 0. \tag{4.14}$$

Observe that for each feasible point $\hat{x} \in \mathbb{R}^3$ *of* (4.14) *we have* $\|\hat{x}\|_0 \le 2$. *We shall now show that every feasible point of* (4.14) *is also an SP-AKKT point. To this end, let* $\hat{x}$ *be feasible for* (4.14). *Then, since* $\|\hat{x}\|_0 \le 2$, $\hat{x}$ *is feasible for* (3.5). *Therefore, it is a CC-AM-stationary point by Example 3.16. By Remark 4.20, it is then an SP-AKKT point of* (4.14). *On the other hand, for a feasible point* $\hat{x} \in \mathbb{R}^3$ *of* (4.14) *to be an SP-CAKKT point, we must have that* $\hat{x}_2 \in \{0, 2\}$: *Suppose that* $\hat{x} \in \mathbb{R}^3$ *is an SP-CAKKT point of* (4.14) *such that* $\hat{x}_2 \notin \{0, 2\}$. *Since* $\|\hat{x}\|_0 \le 2$, *by Remark 4.16* $\hat{x}$ *is then a CC-CAM-stationary point of* (3.5). *This leads to a contradiction.*

By Theorem 4.15 and Theorem 4.21 we obtain the following result.

Theorem 4.23. *Let $\hat{x} \in X$ be a local minimiser of* (1.2). *Then it is an SP-AKKT point of* (1.2).

By Example 4.17 and Theorem 4.21 we know that the converse of Theorem 4.23 is false in general.

4.2 Sequential Constraint Qualifications

Definition 4.19 and Theorem 4.23 naturally lead to the following exact stationarity concept for (1.2).

Definition 4.24. *Let $\hat{x} \in X$. We then say that $\hat{x}$ is an SP-KKT point iff there exist multipliers $\lambda \in \mathbb{R}^m_+$, $\mu \in \mathbb{R}^p$, and $\gamma \in \mathbb{R}^n$ such that*

(a) $0 = \nabla f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i=1}^{n} \gamma_i e_i,$

(b) $\forall i \notin I_g(\hat{x}) : \ \lambda_i = 0,$

(c) $\forall i \in I_\pm(\hat{x}) : \ \gamma_i = 0.$

Remark 4.25. *Note that for each $s \geq \|\hat{x}\|_0$ $\hat{x} \in X$ is an SP-KKT point of* (1.2) *iff it is a CC-M-stationary point of* (1.1).

Due to Remark 4.16, Remark 4.20, and Remark 4.25, most of the ingredients needed to establish the relationships between the sequential optimality conditions introduced in Section 4.1 and SP-KKT can be directly transferred from Section 3.2.

Theorem 4.26 (Theorem 3.20)**.** *Let $\hat{x} \in X$. Then*

$$\hat{x} \text{ is an SP-CAKKT point} \iff -\nabla f(\hat{x}) \in \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)).$$

Theorem 4.27 (Theorem 3.22)**.** *Let $\hat{x} \in X$. Then*

$$\hat{x} \text{ is an SP-KKT point} \iff -\nabla f(\hat{x}) \in K^C((\hat{x},0)).$$

Theorem 4.28 (Theorem 3.23)**.** *Let $\hat{x} \in X$. Then*

$$K^C((\hat{x},0)) \subseteq \limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)).$$

Corollary 4.29 (Corollary 3.24)**.** *Let $\hat{x} \in X$. Then*

$$\hat{x} \text{ is an SP-KKT point} \implies \hat{x} \text{ is an SP-CAKKT point.}$$

The converse is not true in general as the following example shows.

Example 4.30 (Example 3.25)**.** *We know that $(1/2,0)^T$ is the unique global minimiser of* (3.8)*. Let us now fix an arbitrary $\rho > 0$. By Theorem 4.1, $(1/2,0)^T$ is then a local minimiser of the sparse optimisation problem*

$$\min_{x\in\mathbb{R}^2} \ x_1 + 10x_2 + \rho\|x\|_0 \quad \text{s.t.} \quad \left(x_1 - \frac{1}{2}\right)^2 + (x_2-1)^2 \leq 1. \tag{4.15}$$

Hence, by Theorem 4.15, it is then also an SP-CAKKT point of (4.15)*. On the other hand, it was shown in Example 3.25, that $(1/2,0)^T$ cannot be a CC-M-stationary point of* (3.8)*. Thus, by Remark 4.25, it also cannot be an SP-KKT point of* (4.15).

The following is clearly a sufficient condition for the converse of Corollary 4.29 to hold.

Definition 4.31 (Definition 3.26). *A feasible point $\hat{x} \in \mathbb{R}^n$ of* (1.2) *is said to satisfy the SP-CAKKT-regularity condition iff*

$$\limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)) \subseteq K^C((\hat{x},0)).$$

Theorem 4.32 (Theorem 3.27). *Let $\hat{x} \in \mathbb{R}^n$ be an SP-CAKKT point of* (1.2) *which satisfies the SP-CAKKT-regularity condition. Then $\hat{x}$ is an SP-KKT point.*

Example 4.33 (Example 3.28). *If $0 \in \mathbb{R}^n$ is feasible for* (1.2)*, then it is an SP-KKT point and therefore, also an SP-CAKKT point. Moreover, it also satisfies SP-CAKKT-regularity.*

The next theorem states that SP-CAKKT-regularity is a strict constraint qualification with respect to the SP-CAKKT condition.

Theorem 4.34 (Theorem 3.29). *Let $\hat{x} \in X$. Suppose that for every continuously differentiable function $f \in C^1(\mathbb{R}^n, \mathbb{R})$ the following implication holds*

$$\hat{x} \text{ is an SP-CAKKT point } \Longrightarrow \hat{x} \text{ is an SP-KKT point.}$$

Then $\hat{x}$ satisfies SP-CAKKT-regularity.

Theorem 4.35 (Theorem 3.30). *Let $\hat{x} \in X$. Then*

$$\hat{x} \text{ is an SP-AKKT point} \Longleftrightarrow -\nabla f(\hat{x}) \in \limsup_{x\to\hat{x}} K_{\hat{x}}(x).$$

Corollary 4.36 (Corollary 3.33). *Let $\hat{x} \in \mathbb{R}^n$ be feasible for* (1.2)*. Then*

$$\hat{x} \text{ is an SP-KKT point} \Longrightarrow \hat{x} \text{ is an SP-AKKT point.}$$

By Example 4.30 and Theorem 4.21 we know that the converse of Corollary 4.36 is false in general.

Definition 4.37 (Definition 3.34). *A feasible point $\hat{x} \in \mathbb{R}^n$ of* (1.2) *is said to satisfy the SP-AKKT-regularity condition iff*

$$\limsup_{x\to\hat{x}} K_{\hat{x}}(x) \subseteq K_{\hat{x}}(\hat{x}).$$

Theorem 4.38 (Theorem 3.35). *Let $\hat{x} \in \mathbb{R}^n$ be an SP-AKKT point of* (1.2) *which satisfies the SP-AKKT-regularity condition. Then $\hat{x}$ is an SP-KKT point.*

The following theorem states that SP-AKKT-regularity is a strict constraint qualification with respect to the SP-AKKT condition.

Theorem 4.39 (Theorem 3.36). *Let $\hat{x} \in X$. Suppose that for every continuously differentiable function $f \in C^1(\mathbb{R}^n, \mathbb{R})$ the following implication holds*

$$\hat{x} \text{ is an SP-AKKT point } \Longrightarrow \hat{x} \text{ is an SP-KKT point.}$$

Then $\hat{x}$ satisfies SP-AKKT-regularity.

Remark 4.40 (Remark 3.37). *Note that if $0 \in \mathbb{R}^n$ is feasible for* (1.2)*, then it is an SP-AKKT point which satisfies SP-AKKT-regularity.*

Theorem 4.41 (Theorem 3.38). *Let $\hat{x} \in X$. Then*

$$\limsup_{(x,r)\to(\hat{x},0)} K^C((x,r)) \subseteq \limsup_{x\to\hat{x}} K_{\hat{x}}(x).$$

Corollary 4.42 (Corollary 3.39). *Let $\hat{x} \in X$. The following implication then holds*

$$\hat{x} \text{ satisfies SP-AKKT-regularity} \implies \hat{x} \text{ satisfies SP-CAKKT-regularity.}$$

Remark 4.43 (Remark 3.40). *Let $\hat{x} \in X$. Then it is clear that*

- *$\hat{x}$ is an SP-KKT point iff it is a KKT point of* (3.10),
- *$\hat{x}$ is an SP-CAKKT point iff it is a CAKKT point of* (3.10),
- *$\hat{x}$ is an SP-AKKT point iff it is an AKKT point of* (3.10),
- *$\hat{x}$ satisfies SP-CAKKT-regularity iff it satisfies CAKKT-regularity with respect to* (3.10),
- *$\hat{x}$ satisfies SP-AKKT-regularity iff it satisfies AKKT-regularity with respect to* (3.10).

Just like in [23, Definition 3.11], we can now utilise (3.10) to introduce stronger constraint qualifications.

Definition 4.44 (Definition 3.41). *Let $\hat{x} \in X$.Then $\hat{x}$ satisfies*

(a) SP-LICQ iff the gradients

$$\nabla g_i(\hat{x}) \ (i \in I_g(\hat{x})), \ \nabla h_i(\hat{x}) \ (i = 1, \dots, p), \ e_i \ (i \in I_0(\hat{x}))$$

are linearly independent;

(b) SP-MFCQ iff the gradients

$$\nabla g_i(\hat{x}) \ (i \in I_g(\hat{x})) \quad \text{and} \quad \nabla h_i(\hat{x}) \ (i = 1, \dots, p), \ e_i \ (i \in I_0(\hat{x}))$$

are positive-linearly independent;

(c) SP-CPLD iff for any subsets $I_1 \subseteq I_g(\hat{x})$, $I_2 \subseteq \{1, \dots, p\}$, and $I_3 \subseteq I_0(\hat{x})$ such that the gradients

$$\nabla g_i(x) \ (i \in I_1), \quad \text{and} \quad \nabla h_i(x) \ (i \in I_2), \ e_i \ (i \in I_3)$$

are positive-linearly dependent in $x = \hat{x}$, they are linearly dependent in a neighborhood of $\hat{x}$.

Proposition 4.45. *The following relations follow directly from their corresponding NLP relations applied to* (3.10)*:*

$$\text{SP-LICQ} \implies \text{SP-MFCQ} \implies \text{SP-CPLD} \implies \text{SP-AKKT-reg.} \implies \text{SP-CAKKT-reg.}$$

Corollary 4.46 (Corollary 3.43). *If g and h in* (1.1) *are affine-linear, then every feasible point of* (1.2) *satisfies SP-AKKT-regularity, and therefore SP-CAKKT-regularity as well.*

4.3 Relaxed Reformulation

Let us now turn our attention to (1.10), see also [30]. We would like to show that we can solve (1.2) by solving (1.10) instead. First we tackle feasibility.

Theorem 4.47 ([20, Lemma 3.3 (a)]). *Let $\hat{x} \in \mathbb{R}^n$. Then $\hat{x}$ is feasible for* (1.2) *iff there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for* (1.10).

Proof. "$\Longrightarrow$": Suppose that $\hat{x}$ is feasible for (1.2). Then we have $\hat{x} \in X$. Define $\hat{y} \in \mathbb{R}^n$ such that

$$\hat{y}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Clearly we have $\hat{y} \leq e$ and $\hat{x} \circ \hat{y} = 0$. Hence, $(\hat{x}, \hat{y})$ is feasible for (1.10).
"$\Longleftarrow$": Suppose that there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for (1.10). Then this immediately implies that $\hat{x} \in X$ and hence, $\hat{x}$ is feasible for (1.2). □

Observe that for a feasible point $\hat{x} \in \mathbb{R}^n$ of (1.2) the corresponding vector $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for (1.10) is not necessarily unique. Indeed, in the proof of Theorem 4.47, for each $i \in I_0(\hat{x})$ we can replace 1 with any number in $(-\infty, 1]$ and the resulting pair $(\hat{x}, \hat{y})$ would still be feasible for (1.10). Nevertheless, the choice $\hat{y}_i = 1$ for each $i \in I_0(\hat{x})$ is natural since it leads to the smallest objective function value.

Lemma 4.48 ([20, Lemma 3.3 (b)]). *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be feasible for* (1.10). *Then*

$$n - e^T \hat{y} \geq \|\hat{x}\|_0$$

and equality holds iff $\hat{y}_i = 1$ for each $i \in I_0(\hat{x})$.

Proof. The feasibility of $(\hat{x}, \hat{y})$ implies that

(a) $\hat{x} \circ \hat{y} = 0 \implies \hat{y}_i = 0 \ \forall i \in I_\pm(\hat{x})$,

(b) $\hat{y} \leq e \implies 0 \leq 1 - \hat{y}_i \ \forall i \in I_0(\hat{x})$.

Hence,

$$\|\hat{x}\|_0 = \sum_{i \in I_\pm(\hat{x})} 1 \overset{(a)}{=} \sum_{i \in I_\pm(\hat{x})} (1 - \hat{y}_i) \overset{(b)}{\leq} \sum_{i \in I_\pm(\hat{x})} (1 - \hat{y}_i) + \sum_{i \in I_0(\hat{x})} (1 - \hat{y}_i) = n - e^T \hat{y}.$$

Equality obviously holds iff

$$0 = \sum_{i \in I_0(\hat{x})} (1 - \hat{y}_i) \overset{(b)}{\iff} \hat{y}_i = 1 \ \forall i \in I_0(\hat{x}).$$ □

Proposition 4.49 ([20, Lemma 3.4]). *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a local minimiser of* (1.10). *Then $\hat{y}_i = 1$ for each $i \in I_0(\hat{x})$ and therefore, $n - e^T \hat{y} = \|\hat{x}\|_0$.*

Proof. By assumption, $(\hat{x}, \hat{y})$ is a local minimiser of (1.10). Hence, $\hat{y}$ is a global solution of the linear programme

$$\max_y \ e^T y \quad \text{s.t.} \quad y \leq e, \ \hat{x} \circ y = 0.$$

This implies that $\hat{y}_i = 1$ for each $i \in I_0(\hat{x})$ and thus, by Lemma 4.48 we also have $n - e^T \hat{y} = \|\hat{x}\|_0$. □

We can now show that the local solutions of (1.2) and (1.10) are equivalent in a certain sense.

Theorem 4.50 ([20, Theorem 3.5]). *Let $\hat{x} \in \mathbb{R}^n$. Then $\hat{x}$ is a local minimiser of* (1.2) *iff there exists a unique $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a local minimiser of* (1.10). *In this case the vector $\hat{y}$ is given by*

$$\hat{y}_i = \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}) \end{cases}$$

and we have

$$f(\hat{x}) + \rho \|\hat{x}\|_0 = f(\hat{x}) + \rho \left(n - e^T \hat{y}\right).$$

Proof. "$\Longrightarrow$": Assume first that $\hat{x}$ is a local minimiser of (1.2). Then there exists $\epsilon > 0$ such that for each $x \in X$ with $x \in B_\epsilon(\hat{x})$ we have

$$f(\hat{x}) + \rho\|\hat{x}\|_0 \leq f(x) + \rho\|x\|_0. \tag{4.16}$$

Now let $\hat{y}$ be defined as above. By the proof of Theorem 4.47 we know that $(\hat{x}, \hat{y})$ is feasible for (1.10). Suppose now that (x, y) is another feasible point of (1.10) such that $(x, y) \in B_\epsilon((\hat{x}, \hat{y}))$. Then we have

$$(x, y) \in B_\epsilon((\hat{x}, \hat{y})) \Longleftrightarrow \|(x, y) - (\hat{x}, \hat{y})\| < \epsilon \Longrightarrow \|x - \hat{x}\| < \epsilon \Longleftrightarrow x \in B_\epsilon(\hat{x}).$$

Moreover, by Theorem 4.47 we also know that x is feasible for (1.2). Hence,

$$f(\hat{x}) + \rho\left(n - e^T\hat{y}\right) \overset{Lemma\ 4.48}{=} f(\hat{x}) + \rho\|\hat{x}\|_0 \overset{(4.16)}{\leq} f(x) + \rho\|x\|_0 \overset{Lemma\ 4.48}{\leq} f(x) + \rho\left(n - e^T y\right).$$

This implies that $(\hat{x}, \hat{y})$ is a local minimiser of (1.10). Now suppose that there exists another $y \in \mathbb{R}^n$ such that $(\hat{x}, y)$ is also a local minimiser of (1.10). But by Proposition 4.49 it follows that $y = \hat{y}$. This proves the uniqueness of $\hat{y}$.

"$\Longleftarrow$": Now assume that there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a local minimiser of (1.10). By Proposition 4.49 $\hat{y}$ is then uniquely defined as above and we have $\|\hat{x}\|_0 = n - e^T\hat{y}$. Since $(\hat{x}, \hat{y})$ is feasible for (1.10), then by Theorem 4.47 $\hat{x} \in X$. Furthermore, by assumption, there exists $\epsilon_1 > 0$ such that for each feasible point (x, y) of (1.10) with $(x, y) \in B_{\epsilon_1}(\hat{x}, \hat{y})$ we have

$$f(\hat{x}) + \rho\left(n - e^T\hat{y}\right) \leq f(x) + \rho\left(n - e^T y\right). \tag{4.17}$$

Now by Lemma 3.6 there exists $\epsilon_2 > 0$ such that for each $x \in B_{\epsilon_2}(\hat{x})$ we have

$$I_\pm(\hat{x}) \subseteq I_\pm(x).$$

Moreover, by the lower semicontinuity of f there exists $\epsilon_3 > 0$ such that for each $x \in B_{\epsilon_3}(\hat{x})$ we have

$$f(\hat{x}) - f(x) < \rho \quad \Longleftrightarrow \quad f(\hat{x}) < f(x) + \rho.$$

Now define

$$\epsilon := \min_{i=1,2,3} \epsilon_i > 0.$$

Suppose that $x \in X \cap B_\epsilon(\hat{x})$. Since $\epsilon \leq \epsilon_2$ we also have $x \in B_{\epsilon_2}(\hat{x})$. We now differentiate between 2 cases.

<u>Case 1</u>: $I_\pm(\hat{x}) = I_\pm(x)$

In this case we have

$$\|x\|_0 = \mathrm{card}(I_\pm(x)) = \mathrm{card}(I_\pm(\hat{x})) = \|\hat{x}\|_0 = n - e^T\hat{y}.$$

Furthermore, we also have

$$I_0(\hat{x}) = I_0(x).$$

By the structure of $\hat{y}$ and the proof of Theorem 4.47 $(x, \hat{y})$ is then feasible for (1.10). Now since $\epsilon \leq \epsilon_1$ we then have

$$\|(x, \hat{y}) - (\hat{x}, \hat{y})\| = \|x - \hat{x}\| < \epsilon \leq \epsilon_1 \Longrightarrow (x, \hat{y}) \in B_{\epsilon_1}((\hat{x}, \hat{y}))$$

and therefore,

$$f(\hat{x}) + \rho\|\hat{x}\|_0 = f(\hat{x}) + \rho\left(n - e^T\hat{y}\right) \overset{(4.17)}{\leq} f(x) + \rho\left(n - e^T\hat{y}\right) = f(x) + \rho\|x\|_0.$$

<u>Case 2</u>: $I_\pm(\hat{x}) \subsetneq I_\pm(x)$

In this case the index set $I_\pm(x)$ contains at least one more element than $I_\pm(\hat{x})$ and hence,

$$1 \leq \mathrm{card}(I_\pm(x)) - \mathrm{card}(I_\pm(\hat{x})) \quad \Longleftrightarrow \quad \|\hat{x}\|_0 + 1 \leq \|x\|_0.$$

Since $\epsilon \leq \epsilon_3$ we then have

$$f(\hat{x}) < f(x) + \rho \Longrightarrow f(\hat{x}) + \rho\|\hat{x}\|_0 < f(x) + \rho\left(\|\hat{x}\|_0 + 1\right) \leq f(x) + \rho\|x\|_0.$$

Thus, we conclude that $\hat{x}$ is a local minimiser of (1.2). □

Note that the equivalence of global minimisers is a direct consequence of Theorem 4.47 as well as Lemma 4.48 which state the relations of feasible points and objective function values between (1.2) and (1.10). Nevertheless, we state it here for completeness sake.

Theorem 4.51 ([20, Theorem 3.6]). *Let $\hat{x} \in \mathbb{R}^n$. Then $\hat{x}$ is a global minimiser of* (1.2) *iff $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ with*

$$\hat{y}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}) \end{cases}$$

is a global minimiser of (1.10).

In the context of CC, two different exact stationarity concepts were introduced in [21, Definition 4.6]: CC-M-stationarity which corresponds to the KKT-condition of (3.10) and CC-S-stationarity which corresponds to the KKT-condition of the relaxed reformulation. Here these two concepts coincide, hence the single term SP-KKT. We shall now prove this.

Let $z \in \mathbb{R}^n$. Obviously we have $I_1(z) \subseteq I_\pm(z)$. Now a point $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ is said to satisfy the KKT-condition of (1.10) iff it is feasible for (1.10) and there exists $(\lambda, \mu, \tilde{\eta}, \tilde{\gamma}) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n_+ \times \mathbb{R}^n$ such that

(a) $0 = \begin{bmatrix} \nabla f(\hat{x}) \\ -\rho e \end{bmatrix} + \sum_{i=1}^m \lambda_i \begin{bmatrix} \nabla g_i(\hat{x}) \\ 0 \end{bmatrix} + \sum_{i=1}^p \mu_i \begin{bmatrix} \nabla h_i(\hat{x}) \\ 0 \end{bmatrix} + \sum_{i=1}^n \tilde{\eta}_i \begin{bmatrix} 0 \\ e_i \end{bmatrix} + \sum_{i=1}^n \tilde{\gamma}_i \begin{bmatrix} \hat{y}_i e_i \\ \hat{x}_i e_i \end{bmatrix},$

(b) $\lambda_i = 0 \; \forall i \notin I_g(\hat{x})$,

(c) $\tilde{\eta}_i = 0 \; \forall i \notin I_1(\hat{y})$.

Separating the equation with respect to the x- and y-variables, we obtain

$$\begin{aligned} 0 &= \nabla f(\hat{x}) + \sum_{i \in I_g(\hat{x})} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \mu_i \nabla h_i(\hat{x}) + \sum_{i \in I_\pm(\hat{y})} \tilde{\gamma}_i \hat{y}_i e_i, \\ \rho e &= \sum_{i \in I_1(\hat{y})} \tilde{\eta}_i e_i + \sum_{i \in I_\pm(\hat{x})} \tilde{\gamma}_i \hat{x}_i e_i. \end{aligned}$$

The feasibility of $(\hat{x}, \hat{y})$ for (1.10) implies that $I_1(\hat{y}) \cap I_\pm(\hat{x}) = \emptyset$. Clearly $I_1(\hat{y}) \dot\cup I_\pm(\hat{x}) \subseteq \{1, \dots, n\}$. Since $\rho > 0$, the second equation above implies that $\{1, \dots, n\} = I_1(\hat{y}) \dot\cup I_\pm(\hat{x})$ and hence, $I_1(\hat{y}) = I_\pm(\hat{x})^c = I_0(\hat{x})$. This leads to the following.

Lemma 4.52 ([20, Section 5]). *Every KKT-point $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ of* (1.10) *has the property that*

$$\hat{y}_i = \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

The same property is satisfied by all local minimisers of (1.10), cf. Proposition 4.49. Note also that this property of local solutions and KKT-points is very special for sparse optimisation and does, in general, not hold for some of the related classes of minimisation problems.

Theorem 4.53 ([20, Theorem 5.2]). *Let $\hat{x} \in X$. Then $\hat{x}$ is an SP-KKT point of* (1.2) *iff there exists a unique $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a KKT-point of* (1.10)*. The vector $\hat{y}$ is given by*

$$\hat{y}_i = \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Proof. "$\Longrightarrow$": Let $\hat{x} \in X$ be an SP-KKT point of (1.2) with a corresponding multiplier triple $(\lambda, \mu, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$. Define $\hat{y}$ as above. Then $(\hat{x}, \hat{y})$ is feasible for (1.10). Furthermore, we have $I_0(\hat{x}) = I_1(\hat{y}) = I_\pm(\hat{y})$. Define the multipliers

$$\tilde{\gamma}_i := \begin{cases} \frac{\gamma_i}{\hat{y}_i} & \text{if } i \in I_0(\hat{x}), \\ \frac{\rho}{\hat{x}_i} & \text{if } i \in I_\pm(\hat{x}), \end{cases} \quad \text{and} \quad \tilde{\eta}_i := \begin{cases} \rho & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

Then $(\hat{x}, \hat{y})$ together with the multiplier quadruple $(\lambda, \mu, \tilde{\eta}, \tilde{\gamma}) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n_+ \times \mathbb{R}^n$ is a KKT-point of (1.10). Suppose now that $y \in \mathbb{R}^n$ is another vector such that $(\hat{x}, y)$ is a KKT-point of (1.10). Lemma 4.52 then implies that $y = \hat{y}$. This shows the uniqueness.
"$\Longleftarrow$": Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a KKT-point of (1.10) with a corresponding multiplier quadruple $(\lambda, \mu, \tilde{\eta}, \tilde{\gamma}) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n_+ \times \mathbb{R}^n$. By Lemma 4.52, $\hat{y}$ is uniquely defined as above. Furthermore, by Theorem 4.47, $\hat{x}$ is feasible for (1.2). Observe that $I_0(\hat{x}) = I_1(\hat{y}) = I_\pm(\hat{y})$. Define $\gamma \in \mathbb{R}^n$ such that

$$\gamma_i := \begin{cases} \tilde{\gamma}\hat{y}_i & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

Now since $(\hat{x}, \hat{y})$ is a KKT-point we then have $\lambda_i = 0 \ \forall i \notin I_g(\hat{x})$ and

$$\begin{aligned} 0 &= \nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \nabla h_i(\hat{x}) + \sum_{i \in I_\pm(\hat{y})} \tilde{\gamma}_i \hat{y}_i e_i \\ &= \nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \nabla h_i(\hat{x}) + \sum_{i \in I_0(\hat{x})} \gamma_i e_i \\ &= \nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \nabla h_i(\hat{x}) + \sum_{i=1}^n \gamma_i e_i. \end{aligned}$$

Hence, $\hat{x}$ is an SP-KKT point of (1.2) with a corresponding multiplier triple $(\lambda, \mu, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^n$. □

In the context of CC, despite being one of the weakest constraint qualifications for NLP, ACQ is usually violated at a feasible point of the corresponding relaxed programme, see [23, Section 3]. Hence, a CC-tailored ACQ was introduced in [23]. Thus, it is natural to ask if there is a need to introduce SP-tailored ACQ for (1.10) as well. It turns out that this is unnecessary. This is due to Proposition 4.49 which implies that it suffices to only consider those feasible points $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ with $I_0(\hat{x}) \cap I_0(\hat{y}) = \emptyset$. Let us now elaborate this further. To simplify the notation we shall denote the feasible set of (1.10) by Z.

Let $(\hat{x}, \hat{y}) \in Z$. The Bouligand tangent cone, cf. Definition 2.15, of Z at $(\hat{x}, \hat{y})$ is given by

$$T_Z((\hat{x}, \hat{y})) = \left\{ (d_x, d_y) \in \mathbb{R}^n \times \mathbb{R}^n \;\middle|\; \begin{array}{c} \exists \{(x^k, y^k)\} \subseteq Z, \ \{t_k\} \downarrow 0 : \\ \{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y}) \quad \wedge \quad \left\{ \frac{(x^k, y^k) - (\hat{x}, \hat{y})}{t_k} \right\} \longrightarrow (d_x, d_y) \end{array} \right\}$$

and the corresponding linearisation cone, cf. Definition 2.16, is given by

$$L_Z((\hat{x}, \hat{y})) = \left\{ (d_x, d_y) \;\middle|\; \begin{array}{ll} \nabla g_i(\hat{x})^T d_x \leq 0 & \forall i \in I_g(\hat{x}), \\ \nabla h_i(\hat{x})^T d_x = 0 & \forall i = 1, \dots, p, \\ e_i^T d_y \leq 0 & \forall i \in I_1(\hat{y}), \\ \hat{y}_i e_i^T d_x + \hat{x}_i e_i^T d_y = 0 & \forall i = 1, \dots, n \end{array} \right\}.$$

Suppose now that $I_0(\hat{x}) \cap I_0(\hat{y}) = \emptyset$. Then for each $i \in \{1, \dots, n\}$ we have $\hat{x}_i = 0 \dot{\vee} \hat{y}_i = 0$ where $\dot{\vee}$ denotes exclusive or. Since $\{1, \dots, n\} = I_0(\hat{x}) \dot{\cup} I_\pm(\hat{x})$, the linearisation cone can then be further simplified as

$$L_Z((\hat{x}, \hat{y})) = \left\{ (d_x, d_y) \;\middle|\; \begin{array}{ll} \nabla g_i(\hat{x})^T d_x \leq 0 & \forall i \in I_g(\hat{x}), \\ \nabla h_i(\hat{x})^T d_x = 0 & \forall i = 1, \dots, p, \\ e_i^T d_y \leq 0 & \forall i \in I_1(\hat{y}), \\ e_i^T d_y = 0 & \forall i \in I_\pm(\hat{x}), \\ e_i^T d_x = 0 & \forall i \in I_0(\hat{x}) \end{array} \right\}.$$

By [33, Lemma 2.32], the inclusion $T_Z((\hat{x}, \hat{y})) \subseteq L_Z((\hat{x}, \hat{y}))$ always holds. Now by Theorem 4.47 we have $\hat{x} \in X$. Let us denote the feasible set of the corresponding TNLP($\hat{x}$), cf. (3.10), by $X(\hat{x})$. The Bouligand tangent cone of $X(\hat{x})$ at $\hat{x}$ is given by

$$T_{X(\hat{x})}(\hat{x}) = \left\{ d_x \in \mathbb{R}^n \;\middle|\; \exists \{x^k\} \subseteq X(\hat{x}),\ \{t_k\} \downarrow 0 \ :\ \{x^k\} \longrightarrow \hat{x} \wedge \left\{ \frac{x^k - \hat{x}}{t_k} \right\} \longrightarrow d_x \right\}$$

and the corresponding linearisation cone by

$$L_{X(\hat{x})}(\hat{x}) = \left\{ d_x \;\middle|\; \begin{array}{ll} \nabla g_i(\hat{x})^T d_x \leq 0 & \forall i \in I_g(\hat{x}), \\ \nabla h_i(\hat{x})^T d_x = 0 & \forall i = 1, \dots, p, \\ e_i^T d_x = 0 & \forall i \in I_0(\hat{x}) \end{array} \right\}.$$

Again, by [33, Lemma 2.32] we have $T_{X(\hat{x})}(\hat{x}) \subseteq L_{X(\hat{x})}(\hat{x})$.

Theorem 4.54. *Let $(\hat{x}, \hat{y}) \in Z$ such that $I_0(\hat{x}) \cap I_0(\hat{y}) = \emptyset$. Then*

$$L_Z((\hat{x}, \hat{y})) \subseteq T_Z((\hat{x}, \hat{y})) \quad \Longleftrightarrow \quad L_{X(\hat{x})}(\hat{x}) \subseteq T_{X(\hat{x})}(\hat{x}).$$

In other words, ACQ holds for (1.10) *at $(\hat{x}, \hat{y})$ iff ACQ holds for* (3.10) *at $\hat{x}$.*

Proof. "$\Longrightarrow$": Let $d_x \in L_{X(\hat{x})}(\hat{x})$. Define $d_y := 0$. Then $(d_x, d_y) \in L_Z((\hat{x}, \hat{y}))$. By assumption, we then have $(d_x, d_y) \in T_Z((\hat{x}, \hat{y}))$. Hence, in particular, there exist $\{(x^k, y^k)\} \subseteq Z$ and $\{t_k\} \downarrow 0$ such that $\{x^k\} \longrightarrow \hat{x}$, $\{y^k\} \longrightarrow \hat{y}$, and $\left\{ \frac{x^k - \hat{x}}{t_k} \right\} \longrightarrow d_x$. For each $k \in \mathbb{N}$, since $(x^k, y^k) \in Z$, we then have $x^k \in X$. Thus, it remains to show that $x^k \in X(\hat{x}) = \{x \in X \mid x_i = 0\ \forall i \in I_0(\hat{x})\}$. Let $i \in I_0(\hat{x})$. By assumption, we then have $i \notin I_0(\hat{y})$ and hence, $\hat{y}_i \neq 0$. Since $\{y_i^k\} \longrightarrow \hat{y}_i$, we can assume w.l.o.g. that $y_i^k \neq 0\ \forall k \in \mathbb{N}$. Then, since $(x^k, y^k) \in Z$, it follows that $x_i^k = 0\ \forall k \in \mathbb{N}$. Thus, we conclude that $\{x^k\} \subseteq X(\hat{x})$ and therefore, $d_x \in T_{X(\hat{x})}(\hat{x})$.
"$\Longleftarrow$": Let $(d_x, d_y) \in L_Z((\hat{x}, \hat{y}))$. Then $d_x \in L_{X(\hat{x})}(\hat{x})$ and hence, by assumption, $d_x \in T_{X(\hat{x})}(\hat{x})$. By definition, this implies that there exist $\{x^k\} \subseteq X(\hat{x})$, $\{t_k\} \downarrow 0$ such that $\{x^k\} \longrightarrow \hat{x}$ and $\left\{ \frac{x^k - \hat{x}}{t_k} \right\} \longrightarrow d_x$. Let $i \in I_0(\hat{x})$. By assumption we then have $i \in I_\pm(\hat{y})$. Suppose further that $i \in I_\pm(\hat{y}) \setminus I_1(\hat{y})$. Then $\hat{y}_i < 1$ since $(\hat{x}, \hat{y}) \in Z$. Now since $\{t_k\} \downarrow 0$, we can assume w.l.o.g. that for each $k \in \mathbb{N}$ we have $\hat{y}_i + t_k e_i^T d_y \leq 1$. Let us now define for each $k \in \mathbb{N}$ $y^k \in \mathbb{R}^n$ such that

$$y_i^k := \begin{cases} 0 & \text{if } i \in I_0(\hat{y}), \\ \hat{y}_i + t_k e_i^T d_y & \text{if } i \in I_\pm(\hat{y}). \end{cases}$$

By the preceeding discussion, if $i \in I_\pm(\hat{y}) \setminus I_1(\hat{y})$, then $y_i^k \leq 1$. Suppose now that $i \in I_1(\hat{y})$. Since $(d_x, d_y) \in L_Z((\hat{x}, \hat{y}))$, then $e_i^T d_y \leq 0$ and hence, $y_i^k = \hat{y}_i + t_k e_i^T d_y \leq \hat{y}_i = 1$. Thus, we conclude that $y^k \leq e$ for each $k \in \mathbb{N}$. Now let $i \in I_\pm(\hat{x})$. Then we have $i \in I_0(\hat{y})$. Thus, we have for each $k \in \mathbb{N}$ that $y_i^k = 0$ and therefore, $x_i^k y_i^k = 0$. Furthermore, since $x^k \in X(\hat{x})$, then $x_i^k = 0\ \forall i \in I_0(\hat{x})$ and hence, $x_i^k y_i^k = 0$.

Thus, we conclude that $(x^k, y^k) \in Z$ for each $k \in \mathbb{N}$. Since $\{t_k\} \downarrow 0$, by the definition of y^k we clearly have $\{y^k\} \longrightarrow \hat{y}$. Now let $i \in I_\pm(\hat{y})$. Then we have

$$\left\{ \frac{y_i^k - \hat{y}_i}{t_k} \right\} = \left\{ \frac{t_k e_i^T d_y}{t_k} \right\} = \{(d_y)_i\} \longrightarrow (d_y)_i.$$

Suppose now that $i \in I_0(\hat{y})$. By assumption we then have $i \notin I_0(\hat{x})$ and hence, $i \in I_\pm(\hat{x})$. Thus, since $(d_x, d_y) \in L_Z((\hat{x}, \hat{y}))$, it follows that $(d_y)_i = e_i^T d_y = 0$. Hence,

$$\left\{ \frac{y_i^k - \hat{y}_i}{t_k} \right\} = \{0\} \longrightarrow (d_y)_i.$$

This implies that $\left\{ \frac{y^k - \hat{y}}{t_k} \right\} \longrightarrow d_y$. The assertion then follows. □

By Remark 4.43, [8, Theorem 6], and Theorem 4.54 we then have

Corollary 4.55. *Let $(\hat{x}, \hat{y}) \in Z$ such that $I_0(\hat{x}) \cap I_0(\hat{y}) = \emptyset$. Then*

$$\text{SP-CAKKT-regularity holds at } \hat{x} \implies \text{ACQ holds at } (\hat{x}, \hat{y}).$$

Since (1.10) is an instance of (2.1), by Theorem 4.50, Theorem 2.2, and [15, Theorem 3.1], for every local minimiser $\hat{x} \in \mathbb{R}^n$ of (1.2) there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CAKKT and an AKKT point of (1.10). Let us now establish the relationship between the SP-CAKKT condition for (1.2) and the CAKKT condition for (1.10).

Theorem 4.56. *Let $(\hat{x}, \hat{y}) \in Z$. If $(\hat{x}, \hat{y})$ is a CAKKT-point of* (1.10)*, then $\hat{x}$ is an SP-CAKKT point of* (1.2).

We omit the proof since it is similar to the proof of Theorem 3.48. For the converse we have the following result.

Theorem 4.57. *Let $\hat{x} \in X$. If $\hat{x}$ is an SP-CAKKT point, then there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a CAKKT point of* (1.10).

Proof. By assumption there exist sequences $\{x^k\}, \{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, and $\{\mu^k\} \subseteq \mathbb{R}^p$ such that the conditions in Definition 4.14 hold. Observe that for each $i \in \{1, \dots, n\}$ we have

$$\{\gamma_i^k x_i^k\} \longrightarrow 0.$$

This then implies that for each $i \in I_\pm(\hat{x})$ we have

$$\{\gamma_i^k\} \longrightarrow 0.$$

Moreover, for each $i \in I_\pm(\hat{x})$ we can also assume w.l.o.g. that $x_i^k \neq 0 \ \forall k \in \mathbb{N}$. Now define $\hat{y} \in \mathbb{R}^n$ such that

$$\hat{y}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then $(\hat{x}, \hat{y}) \in Z$. Next define for each $k \in \mathbb{N}$ $y^k, \hat{\gamma}^k \in \mathbb{R}^n$ and $\eta^k \in \mathbb{R}^n_+$ such that $y^k := \hat{y}$ and

$$\hat{\gamma}_i^k := \begin{cases} \frac{\rho}{x_i^k} & \text{if } i \in I_\pm(\hat{x}), \\ \gamma_i^k & \text{if } i \in I_0(\hat{x}) \end{cases} \quad \wedge \quad \eta_i^k := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ \rho & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Now if $i \in I_\pm(\hat{x})$ we then obtain

$$\left\{ -\rho + \eta_i^k + \hat{\gamma}_i^k x_i^k \right\} = \left\{ -\rho + \frac{\rho}{x_i^k} x_i^k \right\} = \{0\} \longrightarrow 0.$$

Suppose now that $i \in I_0(\hat{x})$. We then have

$$\left\{-\rho + \eta_i^k + \hat{\gamma}_i^k x_i^k\right\} = \left\{-\rho + \rho + \gamma_i^k x_i^k\right\} = \left\{\gamma_i^k x_i^k\right\} \longrightarrow 0.$$

Hence,

$$\left\{-\rho e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \hat{\gamma}_i^k x_i^k e_i\right\} \longrightarrow 0.$$

Moreover,

$$\begin{aligned}
&\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k y_i^k e_i\right\} \\
&= \left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \gamma_i^k e_i\right\} \\
&= \left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} - \left\{\sum_{i \in I_\pm(\hat{x})} \gamma_i^k e_i\right\} \longrightarrow 0.
\end{aligned}$$

Observe that $\eta_i^k(y_i^k - 1) = 0$ for each $k \in \mathbb{N}$ and for each $i \in \{1, \dots, n\}$. Hence,

$$\begin{aligned}
&\left\{\sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i=1}^n |\eta_i^k(y_i^k - 1)| + \sum_{i=1}^n |\hat{\gamma}_i^k x_i^k y_i^k|\right\} \\
&= \left\{\sum_{i=1}^m |\lambda_i^k g_i(x^k)| + \sum_{i=1}^p |\mu_i^k h_i(x^k)| + \sum_{i \in I_0(\hat{x})} |\gamma_i^k x_i^k|\right\} \longrightarrow 0.
\end{aligned}$$

□

In contrast to Example 3.51, the AKKT condition for (1.10) does imply the SP-AKKT condition for (1.2) as the next theorem shows.

Theorem 4.58. *Let $(\hat{x}, \hat{y}) \in Z$. If $(\hat{x}, \hat{y})$ is an AKKT-point of* (1.10), *then $\hat{x}$ is an SP-AKKT point of* (1.2).

Proof. By assumption there exist sequences $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}_+^m$, $\{\mu^k\} \subseteq \mathbb{R}^p$, $\{\eta^k\} \subseteq \mathbb{R}_+^n$, and $\{\hat{\gamma}^k\} \subseteq \mathbb{R}^n$ such that $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$ and

(a) $\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k y_i^k e_i\right\} \longrightarrow 0,$

(b) $\left\{-\rho e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \hat{\gamma}_i^k x_i^k e_i\right\} \longrightarrow 0,$

(c) $\forall i \notin I_g(\hat{x}) : \ \lambda_i^k = 0 \ \forall k \in \mathbb{N},$

(d) $\forall i \notin I_1(\hat{y}) : \ \eta_i^k = 0 \ \forall k \in \mathbb{N}.$

Suppose that $i \in I_\pm(\hat{x})$. Then $\hat{y}_i = 0$. Hence, since $i \notin I_1(\hat{y})$ we have $\eta_i^k = 0 \ \forall k \in \mathbb{N}$ and therefore,

$$\left\{-\rho + \hat{\gamma}_i^k x_i^k\right\} \longrightarrow 0 \implies \left\{\hat{\gamma}_i^k x_i^k\right\} \longrightarrow \rho \implies \left\{\hat{\gamma}_i^k\right\} \longrightarrow \frac{\rho}{\hat{x}_i}.$$

Now since $\{y_i^k\} \longrightarrow \hat{y}_i = 0$, this implies $\{\hat{\gamma}_i^k y_i^k\} \longrightarrow \frac{\rho}{\hat{x}_i} \cdot 0 = 0$. Define for each $k \in \mathbb{N}$ $\gamma^k \in \mathbb{R}^n$ such that

$$\gamma_i^k := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ \hat{\gamma}_i^k y_i^k & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then

$$\left\{\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^k e_i\right\}$$
$$= \left\{\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i\in I_0(\hat{x})} \hat{\gamma}_i^k y_i^k\right\}$$
$$= \left\{\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \hat{\gamma}_i^k y_i^k\right\} - \left\{\sum_{i\in I_\pm(\hat{x})} \hat{\gamma}_i^k y_i^k\right\} \longrightarrow 0. \qquad \square$$

4.4 Numerical Methods

Just like (1.9), the relaxed reformulation (1.10) enables us to apply methods developed for (2.1) and MPCC to approximate a solution of (1.2). Here we shall concentrate on the augmented Lagrangian method from [4, 15] and the Scholtes regularisation method from [25, 53].

4.4.1 An Augmented Lagrangian Method

Let $\alpha > 0$ be a given penalty parameter. The PHR augmented Lagrangian function for (1.10) is given by

$$L((x,y),\lambda,\mu,\eta,\gamma;\alpha) := f(x) + \rho\left(n - e^T y\right) + \alpha\pi((x,y),\lambda,\mu,\eta,\gamma;\alpha),$$

where $(\lambda,\mu,\eta,\gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}_+^n \times \mathbb{R}^n$ and

$$\pi((x,y),\lambda,\mu,\eta,\gamma;\alpha) := \frac{1}{2}\left\|\left(\left(g(x) + \frac{\lambda}{\alpha}\right)_+, h(x) + \frac{\mu}{\alpha}, \left(y - e + \frac{\eta}{\alpha}\right)_+, x\circ y + \frac{\gamma}{\alpha}\right)\right\|^2$$

is the shifted quadratic penalty term. The algorithm is then stated below.

Algorithm 4.59 (Augmented Lagrangian Method).

(S_0) *Pick* $\lambda_{max} > 0$, $\mu_{min} < \mu_{max}$, $\eta_{max} > 0$, $\gamma_{min} < \gamma_{max}$, $\tau \in (0,1)$, $\sigma > 1$, $\bar{\lambda}^1 \in [0,\lambda_{max}]^m$, $\bar{\mu}^1 \in [\mu_{min},\mu_{max}]^p$, $\bar{\eta}^1 \in [0,\eta_{max}]^n$, $\bar{\gamma}^1 \in [\gamma_{min},\gamma_{max}]^n$, $\alpha_1 > 0$, *and let* $\{\epsilon_k\} \subseteq \mathbb{R}_+$ *such that* $\{\epsilon_k\} \downarrow 0$. *Set* $k \longleftarrow 1$.

(S_1) *Compute* (x^k, y^k) *as an approximate solution of*

$$\min_{x,y} \; L((x,y),\bar{\lambda}^k,\bar{\mu}^k,\bar{\eta}^k,\bar{\gamma}^k;\alpha_k)$$

satisfying

$$\|\nabla_{(x,y)} L((x^k,y^k),\bar{\lambda}^k,\bar{\mu}^k,\bar{\eta}^k,\bar{\gamma}^k;\alpha_k)\| \le \epsilon_k. \tag{4.18}$$

(S_2) *Update the approximate multipliers:*

- $\lambda_i^k := \max\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\} \quad \forall i = 1,\dots,m,$
- $\mu_i^k := \alpha_k h_i(x^k) + \bar{\mu}_i^k \quad \forall i = 1,\dots,p,$
- $\eta_i^k := \max\{0, \alpha_k (y_i^k - 1) + \bar{\eta}_i^k\} \quad \forall i = 1,\dots,n,$
- $\gamma_i^k := \alpha_k x_i^k y_i^k + \bar{\gamma}_i^k \quad \forall i = 1,\dots,n.$

(S_3) *Update the penalty parameter:*
Define

- $U_i^k := \min\left\{-g_i(x^k), \frac{\bar{\lambda}_i^k}{\alpha_k}\right\} \quad \forall i = 1, \dots, m,$
- $W_i^k := \min\left\{-(y_i^k - 1), \frac{\bar{\eta}_i^k}{\alpha_k}\right\} \quad \forall i = 1, \dots, n.$

If $k = 1$ *or*

$$\max\left\{\|U^k\|, \|h(x^k)\|, \|W^k\|, \|x^k \circ y^k\|\right\} \le \tau \max\left\{\begin{matrix}\|U^{k-1}\|, \|h(x^{k-1})\|, \\ \|W^{k-1}\|, \|x^{k-1} \circ y^{k-1}\|\end{matrix}\right\}, \tag{4.19}$$

set $\alpha_{k+1} = \alpha_k$. *Otherwise set* $\alpha_{k+1} = \sigma\alpha_k$.

(S_4) *Update the safeguarded multipliers:*
Compute $\bar{\lambda}^{k+1} \in [0, \lambda_{max}]^m$, $\bar{\mu}^{k+1} \in [\mu_{min}, \mu_{max}]^p$, $\bar{\eta}^{k+1} \in [0, \eta_{max}]^n$, *as well as* $\bar{\gamma}^{k+1} \in [\gamma_{min}, \gamma_{max}]^n$.

(S_5) *Set* $k \longleftarrow k + 1$ *and go to* (S_1).

Theorem 4.60. *Suppose that the sequence* $\{x^k\}$ *generated by Algorithm 4.59 has a limit point* $\hat{x} \in \mathbb{R}^n$, *i.e.,* $\{x^k\}$ *converges on a subsequence to* $\hat{x}$. *Then the corresponding subsequence of* $\{y^k\}$ *is bounded. In particular we can then extract a limit point* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ *of* $\{(x^k, y^k)\}$.

Proof. Let $\hat{x} \in \mathbb{R}^n$ be a limit point of $\{x^k\}$. By passing to a subsequence we can simplify the notation and assume w.l.o.g. that $\{x^k\} \longrightarrow \hat{x}$. Define for each $k \in \mathbb{N}$

$$B^k := \nabla_y L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) = -\rho e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \gamma_i^k x_i^k e_i, \tag{4.20}$$

where the last equality follows from (S_2). By (4.18) and since $\{\epsilon_k\} \downarrow 0$ we know that $\{B^k\} \longrightarrow 0$. Let us now show that $\{y^k\}$ is bounded. First we claim that

$$\forall i \in \{1, \dots, n\} \exists d_i \in \mathbb{R} \forall k \in \mathbb{N} : \ y_i^k \le d_i. \tag{4.21}$$

Suppose not. Then there exists an index $i \in \{1, \dots, n\}$ for which we can construct a subsequence $\{y_i^{k_l}\}$ such that $\{y_i^{k_l}\} \longrightarrow \infty$. We can then assume w.l.o.g. that $y_i^{k_l} > 1$ for each $l \in \mathbb{N}$. Now the convergence of $\{x^k\}$ implies that $\{x_i^k\} \longrightarrow \hat{x}_i$. Observe that by ($S_3$), the sequence of penalty parameters $\{\alpha_k\}$ is nondecreasing. In particular this implies that

$$0 < \alpha_1 \le \alpha_{k_l} \ \forall l \in \mathbb{N}.$$

Hence,

$$\alpha_1(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l} \le \alpha_{k_l}(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l}.$$

Since $\{\bar{\eta}^k\}$ is by definition a bounded sequence, the left hand side tends to ∞. Hence, so must the right hand side as well. We can then assume w.l.o.g. that $\alpha_{k_l}(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l} \ge 0$ for each $l \in \mathbb{N}$ which in turn implies that $\eta_i^{k_l} = \alpha_{k_l}(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l}$, cf. ($S_2$). From (4.20) and ($S_2$) we then obtain

$$\begin{aligned}
\underbrace{B_i^{k_l} + \rho}_{\to \rho} &= \eta_i^{k_l} + \gamma_i^{k_l} x_i^{k_l} \\
&= \alpha_{k_l}(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l} + \underbrace{\alpha_{k_l}}_{>0} \underbrace{\left(x_i^{k_l}\right)^2}_{\ge 0} \underbrace{y_i^{k_l}}_{>0} + \bar{\gamma}_i^{k_l} x_i^{k_l} \\
&\ge \underbrace{\underbrace{\alpha_{k_l}(y_i^{k_l} - 1) + \bar{\eta}_i^{k_l}}_{\to \infty} + \underbrace{\bar{\gamma}_i^{k_l} x_i^{k_l}}_{\text{bounded}}}_{\to \infty}
\end{aligned}$$

which leads to a contradiction. Thus, (4.21) holds. Next we claim that

$$\forall i \in \{1, \dots, n\} \exists c_i \in \mathbb{R} \forall k \in \mathbb{N} : \quad c_i \le y_i^k. \tag{4.22}$$

Suppose not. Then there exists $i \in \{1, \dots, n\}$ for which we can construct a subsequence $\{y_i^{k_l}\}$ which tends to $-\infty$. As such, we can assume w.l.o.g. that $y_i^{k_l} < 0$ for each $l \in \mathbb{N}$. The convergence of $\{x^k\}$ implies that $\{x_i^k\} \longrightarrow \hat{x}_i$. Now since $\alpha_1 \le \alpha_{k_l}$ we then obtain

$$\alpha_{k_l}\left(y_i^{k_l} - 1\right) + \bar{\eta}_i^{k_l} \le \underbrace{\underbrace{\alpha_1\left(y_i^{k_l} - 1\right)}_{\longrightarrow -\infty} + \underbrace{\bar{\eta}_i^{k_l}}_{\text{bounded}}}_{\longrightarrow -\infty}$$

and hence, $\left\{\alpha_{k_l}\left(y_i^{k_l} - 1\right) + \bar{\eta}_i^{k_l}\right\} \longrightarrow -\infty$. Thus, we can assume w.l.o.g. that $\alpha_{k_l}\left(y_i^{k_l} - 1\right) + \bar{\eta}_i^{k_l} < 0$ for each $l \in \mathbb{N}$ which, by (S_2), then implies that $\eta_i^{k_l} = 0 \ \forall l \in \mathbb{N}$. Now from (4.20) we obtain

$$B_i^{k_l} + \rho = \eta_i^{k_l} + \gamma_i^{k_l} x_i^{k_l} = \gamma_i^{k_l} x_i^{k_l} = \alpha_{k_l}(x_i^{k_l})^2 y_i^{k_l} + \bar{\gamma}_i^{k_l} x_i^{k_l}.$$

<u>Case 1</u>: $i \in I_\pm(\hat{x})$
Here we have

$$\begin{aligned}
\alpha_1\left(x_i^{k_l}\right)^2 \le \alpha_{k_l}\left(x_i^{k_l}\right)^2 &\overset{y_i^{k_l}<0}{\Longrightarrow} \alpha_{k_l}\left(x_i^{k_l}\right)^2 y_i^{k_l} \le \alpha_1\left(x_i^{k_l}\right)^2 y_i^{k_l} \\
&\Longrightarrow \underbrace{\alpha_{k_l}\left(x_i^{k_l}\right)^2 y_i^{k_l} + \bar{\gamma}_i^{k_l} x_i^{k_l}}_{=\left\{B_i^{k_l}+\rho\right\}\longrightarrow\rho} \le \underbrace{\underbrace{\alpha_1 \underbrace{\left(x_i^{k_l}\right)^2}_{\longrightarrow \hat{x}_i^2 > 0} \underbrace{y_i^{k_l}}_{\longrightarrow -\infty}}_{\longrightarrow -\infty} + \underbrace{\bar{\gamma}_i^{k_l} x_i^{k_l}}_{\text{bounded}}}_{\longrightarrow -\infty}
\end{aligned}$$

which leads to a contradiction.
<u>Case 2</u>: $i \in I_0(\hat{x})$
Observe that

$$\underbrace{B_i^{k_l} + \rho}_{\longrightarrow\rho} = \underbrace{\alpha_{k_l}}_{>0}\underbrace{\left(x_i^{k_l}\right)^2}_{\ge 0}\underbrace{y_i^{k_l}}_{<0} + \bar{\gamma}_i^{k_l} x_i^{k_l} \le \underbrace{\underbrace{\bar{\gamma}_i^{k_l}}_{\text{bounded}}\ \underbrace{x_i^{k_l}}_{\longrightarrow 0}}_{\longrightarrow 0}.$$

This yields a contradiction since $\rho > 0$. Thus, (4.22) holds. We can now define

$$c := \min_{i=1,\dots,n} c_i \quad \wedge \quad d := \max_{i=1,\dots,n} d_i.$$

Then we have $\{y^k\} \subseteq [c, d]^n$ which implies the boundedness of $\{y^k\}$. The assertion then follows from Bolzano-Weierstraß theorem. ☐

Just like in Section 3.4.1, we shall denote with $\pi_{0,1}((x, y))$ the unshifted quadratic penalty term. Let us now consider the case where the GKL inequality is satisfied by $\pi_{0,1}$ at a feasible limit point $(\hat{x}, \hat{y})$ of Algorithm 4.59. This is for example the case if the nonlinear constraints g_i and h_i are analytic, see the discussion preceeding Theorem 3.61. The following result is a direct consequence of [5, Theorem 5.1] and Theorem 4.56.

Theorem 4.61. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a limit point of the sequence $\{(x^k, y^k)\}$ generated by Algorithm 4.59 that is feasible for* (1.10). *Assume that $\pi_{0,1}$ satisfies the GKL inequality at $(\hat{x}, \hat{y})$. Then $\hat{x}$ is an SP-CAKKT point.*

As a direct consequence of Theorem 4.61 and Theorem 4.32, we obtain the following

Corollary 4.62. *If, in addition to the assumptions in Theorem 4.61, $\hat{x}$ also satisfies SP-CAKKT-regularity, then $\hat{x}$ is an SP-KKT point.*

In the absence of the GKL inequality assumption, the following theorem states that Algorithm 4.59 may still generate an SP-AKKT point. Note that in the theorem we do not make any assumption on $\{y^k\}$.

Theorem 4.63. *Let $\hat{x} \in \mathbb{R}^n$ be a limit point of the sequence $\{x^k\}$ generated by Algorithm 4.59. If $\hat{x}$ is feasible for* (1.2)*, then it is an SP-AKKT point of* (1.2).

Proof. According to Theorem 4.60, there exists a $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a limit point of $\{(x^k, y^k)\}$ generated by Algorithm 4.59. By passing to a subsequence, we can assume w.l.o.g. that $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$. Now define for each $k \in \mathbb{N}$

$$\begin{aligned} A^k &:= \nabla_x L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) \\ &= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k y_i^k e_i, \end{aligned} \tag{4.23}$$

where the last equality follows from (S_2), and let B^k be as in (4.20). By (4.18) and since $\{\epsilon_k\} \downarrow 0$ we know that $\{A^k\} \longrightarrow 0$ and $\{B^k\} \longrightarrow 0$. Observe that by (S_2) we have $\{\lambda^k\} \subseteq \mathbb{R}^m_+$. Furthermore, by (S_3) the sequence of penalty parameters $\{\alpha_k\}$ is nondecreasing. In particular we then have

$$\alpha_k \geq \alpha_1 > 0 \; \forall k \in \mathbb{N}. \tag{4.24}$$

Let us now differentiate between 2 cases.
Case 1: $\{\alpha_k\}$ is bounded.
Observe that by (S_3), the boundedness of $\{\alpha_k\}$ implies that

$$\exists K \in \mathbb{N} \forall k \geq K : \; \alpha_k = \alpha_K.$$

Now let us take a closer look at (S_2). The boundedness of $\{\alpha_k\}$ immediately implies that

- $\forall i \in \{1, \dots, p\} : \; \{\mu_i^k\}$ is bounded,
- $\forall i \in \{1, \dots, n\} : \; \{\gamma_i^k\}$ is bounded.

By passing to subsequences we can assume w.l.o.g. that these sequences converge, i.e.

- $\forall i \in \{1, \dots, p\} \exists \hat{\mu}_i : \; \{\mu_i^k\} \longrightarrow \hat{\mu}_i$,
- $\forall i \in \{1, \dots, n\} \exists \hat{\gamma}_i : \; \{\gamma_i^k\} \longrightarrow \hat{\gamma}_i$.

Now observe that for each $i \in \{1, \dots, m\}$ we have

$$0 \leq \lambda_i^k \leq |\alpha_k g_i(x^k) + \bar{\lambda}_i^k| \quad \forall k \in \mathbb{N}.$$

Thus, $\{\lambda_i^k\}$ is bounded as well and has a convergent subsequence. By passing to subsequences we can assume w.l.o.g. that

$$\forall i \in \{1, \dots, m\} \exists \hat{\lambda}_i : \; \{\lambda_i^k\} \longrightarrow \hat{\lambda}_i \geq 0.$$

Now the boundedness of $\{\alpha_k\}$ and (S_3) also imply that

$$\{\|U^k\|\} \longrightarrow 0 \quad \wedge \quad \{\|x^k \circ y^k\|\} \longrightarrow 0.$$

Let $i \notin I_g(\hat{x})$. By definition, $\{\bar{\lambda}^k\}$ is bounded. Thus, by (4.24) $\left\{\frac{\bar{\lambda}_i^k}{\alpha_k}\right\}$ is bounded as well and therefore has a convergent subsequence. Assume w.l.o.g. that $\left\{\frac{\bar{\lambda}_i^k}{\alpha_k}\right\}$ converges and denote with a_i its limit. We then have

$$0 = \lim_{k\to\infty} \left\|U_i^k\right\| = \|\min\{-g_i(\hat{x}), a_i\}\| \quad \Longrightarrow \quad \min\{-g_i(\hat{x}), a_i\} = 0.$$

Since $-g_i(\hat{x}) > 0$ we then have $a_i = 0$. This then implies that

$$\left\{g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k}\right\} \longrightarrow g_i(\hat{x}) + a_i = g_i(\hat{x}) < 0.$$

Hence, we can assume w.l.o.g. that

$$g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k} < 0 \quad \forall k \in \mathbb{N}.$$

By (4.24) we then obtain

$$\alpha_k g_i(x^k) + \bar{\lambda}_i^k = \alpha_k \left(g_i(x^k) + \frac{\bar{\lambda}_i^k}{\alpha_k}\right) < 0 \quad \forall k \in \mathbb{N}.$$

Thus, by (S_2) we have

$$\lambda_i^k = \max\left\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\right\} = 0 \quad \forall k \in \mathbb{N}. \tag{4.25}$$

As its limit we then have $\hat{\lambda}_i = 0$. Now let $i \in I_\pm(\hat{x})$. Since $\{\|x^k \circ y^k\|\} \longrightarrow 0$, we then have

$$0 = \lim_{k\to\infty} x_i^k y_i^k = \hat{x}_i \hat{y}_i \quad \Longrightarrow \quad \hat{y}_i = 0.$$

By the definition of A^k, letting $k \longrightarrow \infty$ then yields

$$0 = \nabla f(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i=1}^p \hat{\mu}_i \nabla h_i(\hat{x}) + \sum_{i=1}^n \hat{\gamma}_i \hat{y}_i e_i.$$

Since $\hat{\lambda}_i = 0\ \forall i \notin I_g(\hat{x})$ and $\hat{y}_i = 0\ \forall i \in I_\pm(\hat{x})$, we conclude that $\hat{x}$ is an SP-KKT point. By Corollary 4.36 $\hat{x}$ is then an SP-AKKT point of (1.2).

Case 2: $\{\alpha_k\}$ is unbounded.

Since $\{\alpha_k\}$ is nondecreasing, we then have $\{\alpha_k\} \longrightarrow \infty$. Let us show that $\hat{x}$ is an SP-AKKT point. Observe that by (S_2) we have $\{\lambda^k\} \subseteq \mathbb{R}_+^m$. Suppose now that $i \notin I_g(\hat{x})$. Then

$$\alpha_k g_i(x^k) + \bar{\lambda}_i^k \le \underbrace{\underbrace{\alpha_k}_{\to\infty} \underbrace{g_i(x^k)}_{\to g_i(\hat{x})<0} + \lambda_{\max}}_{\to -\infty}.$$

Hence, $\left\{\alpha_k g_i(x^k) + \bar{\lambda}_i^k\right\} \longrightarrow -\infty$ and we can assume w.l.o.g. that $\alpha_k g_i(x^k) + \bar{\lambda}_i^k < 0$ for each $k \in \mathbb{N}$. By (S_2) we then have $\lambda_i^k = \max\{0, \alpha_k g_i(x^k) + \bar{\lambda}_i^k\} = 0$ for each $k \in \mathbb{N}$. Now let $i \in I_\pm(\hat{x})$. We first show that $\hat{y}_i = 0$. By (S_2) $\eta_i^k \ge 0$ for each $k \in \mathbb{N}$. Hence,

$$B_i^k + \rho = \eta_i^k + \gamma_i^k x_i^k \ge \gamma_i^k x_i^k = \left(\alpha_k x_i^k y_i^k + \bar{\gamma}_i^k\right) x_i^k = \alpha_k \left(x_i^k\right)^2 y_i^k + \bar{\gamma}_i^k x_i^k.$$

Since $\left\{\left(x_i^k\right)^2\right\} \longrightarrow (\hat{x}_i)^2 > 0$, we can assume w.l.o.g. that $\left(x_i^k\right)^2 > 0$ for each $k \in \mathbb{N}$. Thus,

$$\underbrace{\underbrace{\frac{1}{\alpha_k}}_{\to 0} \underbrace{\frac{B_i^k + \rho}{(x_i^k)^2}}_{\to \frac{\rho}{(\hat{x}_i)^2}}}_{\to 0} \ge \underbrace{\underbrace{y_i^k}_{\to \hat{y}_i} + \underbrace{\frac{\bar{\gamma}_i^k}{\alpha_k x_i^k}}_{\to 0}}_{\to \hat{y}_i}.$$

Hence, $\hat{y}_i \leq 0$. Then

$$\alpha_k(y_i^k - 1) + \bar{\eta}_i^k \leq \underbrace{\underbrace{\alpha_k}_{\to\infty} \underbrace{y_i^k - 1}_{\to \hat{y}_i - 1 < 0} + \eta_{\max}}_{\to -\infty}.$$

Thus, $\{\alpha_k(y_i^k - 1) + \bar{\eta}_i^k\} \longrightarrow -\infty$ and we can therefore assume w.l.o.g. that $\alpha_k(y_i^k - 1) + \bar{\eta}_i^k < 0 \ \forall k \in \mathbb{N}$ which in turn implies that $\eta_i^k = \max\{0, \alpha_k(y_i^k - 1) + \bar{\eta}_i^k\} = 0 \ \forall k \in \mathbb{N}$. Now since $\left(x_i^k\right)^2 > 0$, then $x_i^k \neq 0$ for each $k \in \mathbb{N}$. Thus,

$$\left\{ \gamma_i^k \right\} = \left\{ \frac{B_i^k + \rho}{x_i^k} \right\} \longrightarrow \frac{\rho}{\hat{x}_i}.$$

Furthermore, by (S_2)

$$\left\{ x_i^k y_i^k \right\} = \left\{ \frac{\gamma_i^k - \bar{\gamma}_i^k}{\alpha_k} \right\} \longrightarrow 0.$$

In particular we have

$$0 = \lim_{k \to \infty} x_i^k y_i^k = \hat{x}_i \hat{y}_i \quad \Longrightarrow \quad \hat{y}_i = 0.$$

This implies that

$$\left\{ \gamma_i^k y_i^k \right\} \longrightarrow \frac{\rho}{\hat{x}_i} \hat{y}_i = 0.$$

Now define for each $i \in \{1, \dots, n\}$ and for each $k \in \mathbb{N}$

$$\hat{\gamma}_i^k := \begin{cases} \gamma_i^k y_i^k & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

Then,

$$\underbrace{A^k - \sum_{i \in I_\pm(\hat{x})} \gamma_i^k y_i^k e_i}_{\to 0} = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k e_i.$$

Thus, $\hat{x}$ is an SP-AKKT point. □

Then as an immediate consequence of Theorem 4.63 and Theorem 4.38, we have the following result.

Corollary 4.64. *If, in addition to the assumptions in Theorem 4.63, $\hat{x}$ also satisfies SP-AKKT-regularity, then $\hat{x}$ is an SP-KKT point.*

Theorem 4.61 and Theorem 4.63 assume that the limit point $\hat{x}$ is feasible for (1.2). The next theorem shows that this assumption is plausible.

Theorem 4.65. *Let $\hat{x} \in \mathbb{R}^n$ be a limit point of the sequence $\{x^k\}$ generated by Algorithm 4.59. Then $\hat{x}$ is a stationary point of*

$$\min_{x \in \mathbb{R}^n} \|h(x)\|^2 + \|g(x)_+\|^2. \tag{4.26}$$

Proof. We differentiate between 2 cases.
Case 1: $\{\alpha_k\}$ is bounded.
By the same argument as in the proof of Theorem 4.63, the boundedness of $\{\alpha_k\}$ implies that

$$\{\|h(x^k)\|\} \longrightarrow 0 \quad \wedge \quad \{\|U^k\|\} \longrightarrow 0.$$

We then have $\|h(\hat{x})\| = 0$. Moreover, just like in Theorem 4.63 for each $i \in \{1, \dots, m\}$ we can assume w.l.o.g. that there exists $a_i \in \mathbb{R}$ such that $\left\{ \frac{\bar{\lambda}_i^k}{\alpha_k} \right\} \longrightarrow a_i$. Thus, by the definition of U_i^k we then have $\min\{-g_i(\hat{x}), a_i\} = 0$ which implies that

$$g_i(\hat{x}) \leq 0 \quad \Longleftrightarrow \quad g_i(\hat{x})_+ = 0.$$

Hence, $\|g(\hat{x})_+\| = 0$. Consequently $\hat{x}$ is a global minimiser of (4.26) and therefore also a stationary point of the problem.
Case 2: $\{\alpha_k\}$ is unbounded
By passing to a subsequence we can assume w.l.o.g. that $\{\alpha_k\} \longrightarrow \infty$. Furthermore, by 4.60 there exists $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is a limit point of $\{(x^k, y^k)\}$. Assume w.l.o.g. that $\{(x^k, y^k)\} \longrightarrow (\hat{x}, \hat{y})$. Define for each $k \in \mathbb{N}$ A^k as in (4.23) and B^k as in (4.20). Then $\{A^k\} \longrightarrow 0$ and $\{B^k\} \longrightarrow 0$. By (S_2) we have

$$0 = \lim_{k\to\infty} \frac{A^k}{\alpha_k} = \sum_{i=1}^{m} \max\{0, g_i(\hat{x})\}\nabla g_i(\hat{x}) + \sum_{i=1}^{p} h_i(\hat{x})\nabla h_i(\hat{x}) + \sum_{i=1}^{n} \hat{x}_i\,(\hat{y}_i)^2\, e_i.$$

Now if $i \in I_\pm(\hat{x})$, then using the same argument as in the proof of Theorem 4.63 we know that $\hat{y}_i = 0$. Thus, $\hat{x}_i\,(\hat{y}_i)^2 = 0 \ \forall i = 1, \dots, n$. This then implies that

$$0 = \sum_{i=1}^{m} \max\{0, g_i(\hat{x})\}\nabla g_i(\hat{x}) + \sum_{i=1}^{p} h_i(\hat{x})\nabla h_i(\hat{x}) \quad \Longleftrightarrow \quad \nabla\left(\|h(\hat{x})\|^2 + \|g(\hat{x})_+\|^2\right) = 0.$$

Hence, $\hat{x}$ is a stationary point of (4.26). □

Let us now justify (S_1), i.e. we would like to investigate under what conditions the subproblems admit solutions. Our analysis follows [17, Theorem 3.3] which employs exterior penalty method, cf. [16]. This approach was also suggested in [31, page 393]. For the remaining of this subsection we shall assume that f, g, and h are twice continuously differentiable.

Let $\hat{x} \in X$. We define the SP-linearisation cone of X at $\hat{x}$ as

$$L_X(\hat{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla g_i(\hat{x})^T d \leq 0 & \forall i \in I_g(\hat{x}), \\ \nabla h_i(\hat{x})^T d = 0 & \forall i = 1, \dots, p, \\ e_i^T d = 0 & \forall i \in I_0(\hat{x}) \end{array} \right\}.$$

Suppose now that $(\lambda, \mu, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ such that $(\hat{x}, \lambda, \mu, \gamma)$ is an SP-KKT tuple of (1.2). We define

$$I_g^+((\hat{x}, \lambda)) := \{i \in I_g(\hat{x}) \mid \lambda_i > 0\} \quad \wedge \quad I_g^0((\hat{x}, \lambda)) := \{i \in I_g(\hat{x}) \mid \lambda_i = 0\}.$$

Clearly we have $I_g(\hat{x}) = I_g^+((\hat{x}, \lambda)) \,\dot{\cup}\, I_g^0((\hat{x}, \lambda))$. We define the SP-critical cone as

$$C_X((\hat{x}, \lambda, \mu, \gamma)) := \{d \in L_X(\hat{x}) \mid \nabla g_i(\hat{x})^T d = 0 \ \forall i \in I_g^+((\hat{x}, \lambda))\}.$$

Definition 4.66. *An SP-KKT tuple $(\hat{x}, \lambda, \mu, \gamma) \in X \times \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ of (1.2) is said to satisfy the SP second order sufficient condition (SP-SOSC) iff*

$$d^T \left(\nabla^2 f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(\hat{x}) \right) d > 0 \quad \forall d \in C_X((\hat{x}, \lambda, \mu, \gamma)) \setminus \{0\}.$$

Theorem 4.67. *Let $(\hat{x}, \lambda, \mu, \gamma) \in X \times \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ be an SP-KKT tuple of (1.2) which satisfies SP-SOSC. Then $\hat{x}$ satisfies the quadratic growth condition, i.e.*

$$\exists c, \epsilon > 0 \forall x \in B_\epsilon(\hat{x}) \cap X \,:\; f(x) + \rho\|x\|_0 \geq f(\hat{x}) + \rho\|\hat{x}\|_0 + c\|x - \hat{x}\|^2.$$

This then implies that $\hat{x}$ is a strict local minimiser of (1.2).

Proof. Suppose not. Then we can construct a sequence $\{x^k\} \subseteq X$ such that $\{x^k\} \longrightarrow \hat{x}$ and

$$f(x^k) + \rho\|x^k\|_0 < f(\hat{x}) + \rho\|\hat{x}\|_0 + \frac{1}{k}\|x^k - \hat{x}\|^2 \quad \forall k \in \mathbb{N}.$$

This then implies that $x^k \neq \hat{x}\ \forall k \in \mathbb{N}$ and therefore, in particular, $\|x^k - \hat{x}\| > 0\ \forall k \in \mathbb{N}$. Hence, $\left\{\frac{x^k-\hat{x}}{\|x^k-\hat{x}\|}\right\}$ is well-defined and bounded with length 1. By passing to a subsequence we can assume w.l.o.g. that it converges, i.e. $\exists d \in \mathbb{R}^n \setminus \{0\}\ :\ \left\{\frac{x^k-\hat{x}}{\|x^k-\hat{x}\|}\right\} \longrightarrow d$. Furthermore, by Lemma 3.6 we can assume w.l.o.g. that $\|x^k\|_0 \geq \|\hat{x}\|_0\ \forall k \in \mathbb{N}$. Thus,

$$f(x^k) \leq f(x^k) + \rho\underbrace{\left(\|x^k\|_0 - \|\hat{x}\|_0\right)}_{\geq 0} < f(\hat{x}) + \frac{1}{k}\|x^k - \hat{x}\|^2 \quad \forall k \in \mathbb{N}.$$

Now by the second order Taylor expansion there exists for each $k \in \mathbb{N}$ a $\xi^k \in [\![x^k, \hat{x}]\!]$ such that

$$\frac{1}{k}\|x^k - \hat{x}\|^2 > f(x^k) - f(\hat{x}) = \nabla f(\hat{x})^T(x^k - \hat{x}) + \frac{1}{2}(x^k - \hat{x})^T\nabla^2 f(\xi^k)(x^k - \hat{x}). \tag{4.27}$$

By assumption $(\hat{x}, \lambda, \mu, \gamma)$ is an SP-KKT tuple. Hence,

$$\nabla f(\hat{x}) = -\left(\sum_{i\in I_g(\hat{x})} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i\in I_0(\hat{x})} \gamma_i e_i\right).$$

Thus, by (4.27)

$$\begin{aligned}\frac{1}{k}\|x^k - \hat{x}\|^2 > &-\left(\sum_{i\in I_g(\hat{x})} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\hat{x}) + \sum_{i\in I_0(\hat{x})} \gamma_i e_i\right)^T (x^k - \hat{x}) \\ &+ \frac{1}{2}(x^k - \hat{x})^T\nabla^2 f(\xi^k)(x^k - \hat{x}).\end{aligned} \tag{4.28}$$

Now also by the second order Taylor expansion we have

- $\forall i \in I_g(\hat{x}) \forall k \in \mathbb{N}\ \exists \zeta^{i,k} \in [\![x^k, \hat{x}]\!]$ such that

$$\begin{aligned}0 \overset{x^k\in X}{\geq} g_i(x^k) &= g_i(\hat{x}) + \nabla g_i(\hat{x})^T(x^k - \hat{x}) + \frac{1}{2}(x^k - \hat{x})^T\nabla^2 g_i(\zeta^{i,k})(x^k - \hat{x}) \\ &\overset{g_i(\hat{x})=0}{=} \nabla g_i(\hat{x})^T(x^k - \hat{x}) + \frac{1}{2}(x^k - \hat{x})^T\nabla^2 g_i(\zeta^{i,k})(x^k - \hat{x})\end{aligned}$$

  and hence, since $\lambda_i \geq 0$

$$-\lambda_i \nabla g_i(\hat{x})^T(x^k - \hat{x}) \geq \frac{\lambda_i}{2}(x^k - \hat{x})^T\nabla^2 g_i(\zeta^{i,k})(x^k - \hat{x}).$$

- $\forall i \in \{1, \dots, p\} \forall k \in \mathbb{N}\ \exists \eta^{i,k} \in [\![x^k, \hat{x}]\!]$ such that

$$\begin{aligned}0 \overset{x^k\in X}{=} h_i(x^k) &= h_i(\hat{x}) + \nabla h_i(\hat{x})^T(x^k - \hat{x}) + \frac{1}{2}(x^k - \hat{x})^T\nabla^2 h_i(\eta^{i,k})(x^k - \hat{x}) \\ &\overset{h_i(\hat{x})=0}{=} \nabla h_i(\hat{x})^T(x^k - \hat{x}) + \frac{1}{2}(x^k - \hat{x})^T\nabla^2 h_i(\eta^{i,k})(x^k - \hat{x}).\end{aligned}$$

  Thus,

$$-\mu_i \nabla h_i(\hat{x})^T(x^k - \hat{x}) = \frac{\mu_i}{2}(x^k - \hat{x})^T\nabla^2 h_i(\eta^{i,k})(x^k - \hat{x}).$$

Now let $i \in I_0(\hat{x})$. We claim that we can construct a subsequence of $\{x^k\}$ such that the i-th component is equal to 0. Suppose not. Then we can assume w.l.o.g. that $x_i^k \neq 0 \;\forall k \in \mathbb{N}$. Hence, $i \in I_\pm(x^k) \setminus I_\pm(\hat{x}) \;\forall k \in \mathbb{N}$. By Lemma 3.6 we then have $\|x^k\|_0 \geq \|\hat{x}\|_0 + 1$. Shrinking $\epsilon > 0$ from Lemma 3.6 and passing to a subsequence if necessary, by the lower semicontinuity of f we also have $f(\hat{x}) < f(x^k) + \frac{\rho}{2} \;\forall k \in \mathbb{N}$. Furthermore, since $\left\{\frac{1}{k}\|x^k - \hat{x}\|^2\right\} \longrightarrow 0$, by passing to a subsequence we can assume w.l.o.g. that $\frac{1}{k}\|x^k - \hat{x}\|^2 < \frac{\rho}{2} \;\forall k \in \mathbb{N}$. Hence, we have

$$f(\hat{x}) + \frac{1}{k}\|x^k - \hat{x}\|^2 < f(x^k) + \rho \implies f(\hat{x}) + \rho\|\hat{x}\|_0 + \frac{1}{k}\|x^k - \hat{x}\|^2 < f(x^k) + \rho(1 + \|\hat{x}\|_0) \leq f(x^k) + \rho\|x^k\|_0,$$

which leads to a contradiction. Thus, by passing to a subsequence, we can assume w.l.o.g. that $x_i^k = 0 \;\forall k \in \mathbb{N}$. This implies that

$$e_i^T(x^k - \hat{x}) = x_i^k - \hat{x}_i = 0 \;\forall k \in \mathbb{N}.$$

With these considerations we obtain from (4.28)

$$\frac{1}{k}\|x^k - \hat{x}\|^2 > \frac{1}{2}(x^k - \hat{x})^T \left(\nabla^2 f(\xi^k) + \sum_{i \in I_g(\hat{x})} \lambda_i \nabla^2 g_i(\zeta^{i,k}) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(\eta^{i,k}) \right)(x^k - \hat{x})$$

and hence,

$$\frac{2}{k} > \left(\frac{x^k - \hat{x}}{\|x^k - \hat{x}\|} \right)^T \left(\nabla^2 f(\xi^k) + \sum_{i \in I_g(\hat{x})} \lambda_i \nabla^2 g_i(\zeta^{i,k}) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(\eta^{i,k}) \right) \left(\frac{x^k - \hat{x}}{\|x^k - \hat{x}\|} \right).$$

Letting $k \longrightarrow \infty$ then yields

$$\begin{aligned} 0 &\geq d^T \left(\nabla^2 f(\hat{x}) + \sum_{i \in I_g(\hat{x})} \lambda_i \nabla^2 g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(\hat{x}) \right) d \\ &= d^T \left(\nabla^2 f(\hat{x}) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(\hat{x}) \right) d. \end{aligned}$$

To arrive at the desired contradiction, it remains to prove that $d \in C_X((\hat{x}, \lambda, \mu, \gamma))$.

Let $i \in I_g(\hat{x})$. Then by mean value theorem there exists for each $k \in \mathbb{N}$ $y^{i,k} \in [\![x^k, \hat{x}]\!]$ such that $0 \overset{x^k \in X}{\geq} g_i(x^k) = g_i(x^k) - g_i(\hat{x}) = \nabla g_i(y^{i,k})^T(x^k - \hat{x})$. Thus, $\nabla g_i(y^{i,k})^T \frac{x^k - \hat{x}}{\|x^k - \hat{x}\|} \leq 0$. Letting $k \longrightarrow \infty$ yields $\nabla g_i(\hat{x})^T d \leq 0$.

Let $i \in \{1, \dots, p\}$. By mean value theorem there exists for each $k \in \mathbb{N}$ $z^{i,k} \in [\![x^k, \hat{x}]\!]$ such that $0 \overset{x^k, \hat{x} \in X}{=} h_i(x^k) - h_i(\hat{x}) = \nabla h_i(z^{i,k})^T(x^k - \hat{x})$ and hence, $0 = \nabla h_i(z^{i,k})^T \frac{x^k - \hat{x}}{\|x^k - \hat{x}\|}$. Letting $k \longrightarrow \infty$ yields $\nabla h_i(\hat{x})^T d = 0$.

Let $i \in I_0(\hat{x})$. We have already shown that $x_i^k = 0 \;\forall k \in \mathbb{N}$. Hence, $0 = \frac{x_i^k - \hat{x}_i}{\|x^k - \hat{x}\|} = e_i^T \frac{x^k - \hat{x}}{\|x^k - \hat{x}\|}$. Thus, letting $k \longrightarrow \infty$ yields $e_i^T d = 0$.

By mean value theorem for each $k \in \mathbb{N}$ there exists $\theta^k \in [\![x^k, \hat{x}]\!]$ such that $\frac{1}{k}\|x^k - \hat{x}\|^2 > f(x^k) - f(\hat{x}) = \nabla f(\theta^k)^T(x^k - \hat{x})$. Thus, $\frac{1}{k}\|x^k - \hat{x}\| > \nabla f(\theta^k)^T \frac{x^k - \hat{x}}{\|x^k - \hat{x}\|}$. Letting $k \longrightarrow \infty$ yields $0 \geq \nabla f(\hat{x})^T d$. We have shown that $\nabla g_i(\hat{x})^T d \leq 0 \;\forall i \in I_g(\hat{x})$. Now we claim that $\nabla g_i(\hat{x})^T d = 0 \;\forall i \in I_g^+((\hat{x}, \lambda))$. Suppose not. Then $\exists j \in I_g^+((\hat{x}, \lambda)) : \nabla g_j(\hat{x})^T d < 0$. Since $(\hat{x}, \lambda, \mu, \gamma)$ is an SP-KKT tuple we obtain

$$\begin{aligned} 0 &= \underbrace{\nabla f(\hat{x})^T d}_{\leq 0} + \sum_{i \in I_g^+((\hat{x}, \lambda))} \lambda_i \nabla g_i(\hat{x}) + \sum_{i=1}^{p} \mu_i \underbrace{\nabla h_i(\hat{x})^T d}_{=0} + \sum_{i \in I_0(\hat{x})} \gamma_i \underbrace{e_i^T d}_{=0} \\ &\leq \sum_{i \in I_g^+((\hat{x}, \lambda))} \lambda_i \nabla g_i(\hat{x})^T d \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i \in I_g^+((\hat{x},\lambda))\setminus\{j\}} \lambda_i \nabla g_i(\hat{x})^T d + \underbrace{\lambda_j}_{>0} \underbrace{\nabla g_j(\hat{x})^T d}_{<0} \\
&< \sum_{i \in I_g^+((\hat{x},\lambda))\setminus\{j\}} \underbrace{\lambda_i}_{>0} \underbrace{\nabla g_i(\hat{x})^T d}_{\leq 0} \\
&\leq 0,
\end{aligned}
$$

a contradiction. Hence, $\nabla g_i(\hat{x})^T d = 0 \ \forall i \in I_g^+((\hat{x}, \lambda))$ and therefore, $d \in C_X((\hat{x}, \lambda, \mu, \gamma))$. This leads to the desired contradiction and we conclude that the quadratic growth condition holds at $\hat{x}$. This further implies that $\hat{x}$ is a strict local minimiser of (1.2). □

To prove the next result we follow [17, Theorem 3.3]. Some parts of the proof are adapted from [16, Theorem 1, Theorem 2], see also [15, Theorem 5.1, Theorem 5.2]

Theorem 4.68. *Let* $(\hat{x}, \lambda, \mu, \gamma) \in X \times \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^n$ *be an SP-KKT tuple of* (1.2) *which satisfies SP-SOSC. Then there exists* $\hat{\alpha} > 0$ *such that if* $\alpha_1 \geq \hat{\alpha}$ *then there exists a sequence* $\{(x^k, y^k)\}$ *generated by Algorithm 4.59 such that* $\{x^k\} \longrightarrow \hat{x}$.

Proof. By Theorem 4.67 there exist $c, \epsilon > 0$ such that

$$f(\hat{x}) + \rho\|\hat{x}\|_0 + c\|x - \hat{x}\|^2 \leq f(x) + \rho\|x\|_0 \quad \forall x \in B_\epsilon(\hat{x}) \cap X.$$

Hence, $\hat{x}$ is the unique global minimiser of

$$\min_{x \in \mathbb{R}^n} f(x) + \rho\|x\|_0 \quad \text{s.t.} \quad x \in \bar{B}_{\frac{\epsilon}{2}} \cap X. \tag{4.29}$$

Define $\hat{y} \in \mathbb{R}^n$ such that

$$\hat{y}_i := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ 1 & \text{if } i \in I_0(\hat{x}). \end{cases}$$

We claim that $(\hat{x}, \hat{y})$ is then the unique global minimiser of

$$\min_{x,y \in \mathbb{R}^n} f(x) + \rho(n - e^T y) \quad \text{s.t.} \quad (x, y) \in \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y})) \cap Z \tag{4.30}$$

where Z denotes the feasible set of (1.10). Since $\hat{x} \in X$, by the proof of Theorem 4.47 we know that $(\hat{x}, \hat{y}) \in Z$ and hence, $(\hat{x}, \hat{y})$ is feasible for (4.30). Now let (x, y) be another feasible point of (4.30) such that $(x, y) \neq (\hat{x}, \hat{y})$. Observe that since $(x, y) \in Z$, Theorem 4.47 then implies that $x \in X$. Moreover, since $(x, y) \in \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$, then $x \in \bar{B}_{\frac{\epsilon}{2}}(\hat{x})$. Thus, x is feasible for (4.29). Now since $(x, y) \neq (\hat{x}, \hat{y})$, then we have $x \neq \hat{x} \ \vee \ y \neq \hat{y}$. We differentiate now between 2 cases.
Case 1: $x \neq \hat{x}$
We have shown that x is feasible for (4.29). Now by Lemma 4.48 we then have

$$f(\hat{x}) + \rho(n - e^T \hat{y}) = f(\hat{x}) + \rho\|\hat{x}\|_0 < f(x) + \rho\|x\|_0 \leq f(x) + \rho(x - e^T y).$$

Case 2: $x = \hat{x}$
Then we have $y = \hat{y}$. Let us now investigate on which components they can differ. Observe that since $x = \hat{x}$ then we have $I_\pm(x) = I_\pm(\hat{x})$. Thus, by the feasibility of (x, y) and $(\hat{x}, \hat{y})$ for (1.10) we obtain for each $i \in I_\pm(\hat{x})$ that $\hat{y}_i = 0 = y_i$. We can therefore conclude that y and $\hat{y}$ differ on $I_0(\hat{x})$, i.e. there exists $j \in I_0(\hat{x})$ such that $\hat{y}_j \neq y_j$. By definition we have $\hat{y}_j = 1$. On the other hand, the fact that $(x, y) \in Z$ implies that $y_i \leq 1$ for each $i \in \{1, \ldots, n\}$. In particular, by the definition of $\hat{y}$, this means that $y_i \leq 1 = \hat{y}_i \ \forall i \in I_0(\hat{x})$. Now special for j we then must have $y_j < 1 = \hat{y}_j$. Summing over $I_0(\hat{x})$ then yields $\sum_{i \in I_0(\hat{x})} y_i \overset{y_j < \hat{y}_j}{<} \sum_{i \in I_0(\hat{x})} \hat{y}_i$ and therefore, since $y_i = \hat{y}_i = 0 \ \forall i \in I_\pm(\hat{x})$

$$\sum_{i=1}^n y_i < \sum_{i=1}^n \hat{y}_i \quad \Longleftrightarrow \quad e^T y < e^T \hat{y} \quad \Longleftrightarrow \quad n - e^T \hat{y} < n - e^T y.$$

As such

$$f(x) + \rho(n - e^T y) > f(x) + \rho(n - e^T \hat{y}) \overset{x=\hat{x}}{=} f(\hat{x}) + \rho(n - e^T \hat{y}).$$

We can now conclude that $(\hat{x}, \hat{y})$ is the unique global minimiser of (4.30).

Shrinking ϵ if necessary, by the continuity of g we can assume w.l.o.g. that $g_i(x) < 0$ for each $x \in B_\epsilon(\hat{x})$ and for each $i \notin I_g(\hat{x})$. Moreover, we can also assume w.l.o.g. that for each $i \in I_\pm(\hat{x})$ we have $y_i < 1$ for each $y \in B_\epsilon(\hat{y})$. Now let $\alpha > 0$, $\bar{\lambda} \in [0, \lambda_{\max}]^m$, $\bar{\mu} \in [\mu_{\min}, \mu_{\max}]^p$, $\bar{\eta} \in [0, \eta_{\max}]^n$, and $\bar{\gamma} \in [\gamma_{\min}, \gamma_{\max}]^n$. Consider

$$\min_{x,y \in \mathbb{R}^n} L((x,y), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha) \quad \text{s.t.} \quad (x,y) \in \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y})). \tag{4.31}$$

The PHR augmented Lagrangian function is clearly continuous. Moreover, the constraint set is compact. Hence, (4.31) admits a global minimiser $(x,y)(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}, \alpha)$. Now we claim that there exists $\bar{\alpha} > 0$ such that for each $\alpha \geq \bar{\alpha}$ we have

$$(x,y)(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}, \alpha) \in B_{\frac{\epsilon}{2}}((\hat{x}, \hat{y})). \tag{4.32}$$

Suppose not. Then we can construct a sequence of penalty parameters $\{\alpha_k\}$ which tends to ∞ such that for each $k \in \mathbb{N}$ we have for the corresponding global minimiser $(u^k, w^k) := (x^k, y^k)(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}, \alpha)$ of (4.31) that $(u^k, w^k) \in \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y})) \setminus B_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$. This implies that $\|(u^k, w^k) - (\hat{x}, \hat{y})\| = \frac{\epsilon}{2}$. Now since $\{(u^k, w^k)\}$ is a bounded sequence, by passing to a subsequence we can assume w.l.o.g. that it converges, i.e. there exists (u, w) such that $\{(u^k, w^k)\} \longrightarrow (u, w)$. By the property of (u^k, w^k) we then must have $\|(u, w) - (\hat{x}, \hat{y})\| = \frac{\epsilon}{2}$. This implies that $(u, w) \neq (\hat{x}, \hat{y})$. Now observe that since (u^k, w^k) is a global minimiser of (4.31) and $(\hat{x}, \hat{y})$ is also feasible for the problem, we then have

$$L((u^k, w^k), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) \leq L((\hat{x}, \hat{y}), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k). \tag{4.33}$$

Recall that $(\hat{x}, \hat{y}) \in Z$. Let $i \in \{1, \dots, m\}$. Then

$$g_i(\hat{x}) \leq 0 \Longrightarrow g_i(\hat{x}) + \frac{\bar{\lambda}_i}{\alpha_k} \leq \frac{\bar{\lambda}_i}{\alpha_k} \overset{\frac{\bar{\lambda}_i}{\alpha_k} \geq 0}{\Longrightarrow} 0 \leq \max\left\{0, g_i(\hat{x}) + \frac{\bar{\lambda}_i}{\alpha_k}\right\} \leq \frac{\bar{\lambda}_i}{\alpha_k} \Longrightarrow \left(g_i(\hat{x}) + \frac{\bar{\lambda}_i}{\alpha_k}\right)_+^2 \leq \left(\frac{\bar{\lambda}_i}{\alpha_k}\right)^2.$$

This implies that $\left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 \leq \left\|\frac{\bar{\lambda}}{\alpha_k}\right\|^2$. Now let $i \in \{1, \dots, p\}$. We have

$$h_i(\hat{x}) = 0 \Longrightarrow h_i(\hat{x}) + \frac{\bar{\mu}_i}{\alpha_k} = \frac{\bar{\mu}_i}{\alpha_k} \Longrightarrow \left(h_i(\hat{x}) + \frac{\bar{\mu}_i}{\alpha_k}\right)^2 = \left(\frac{\bar{\mu}_i}{\alpha_k}\right)^2.$$

Thus, $\left\|h(\hat{x}) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 = \left\|\frac{\bar{\mu}}{\alpha_k}\right\|^2$. In similar fashion we obtain that $\left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 \leq \left\|\frac{\bar{\eta}}{\alpha_k}\right\|^2$ and $\left\|\hat{x} \circ \hat{y} + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 = \left\|\frac{\bar{\gamma}}{\alpha_k}\right\|^2$. As such,

$$\begin{aligned} L((\hat{x}, \hat{y}), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) &= f(\hat{x}) + \rho\left(n - e^T \hat{y}\right) + \frac{\alpha_k}{2}\begin{pmatrix} \left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(\hat{x}) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|\hat{x} \circ \hat{y} + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{pmatrix} \\ &\leq f(\hat{x}) + \rho\left(n - e^T \hat{y}\right) + \frac{\alpha_k}{2}\left(\frac{\|\bar{\lambda}\|^2}{\alpha_k^2} + \frac{\|\bar{\mu}\|^2}{\alpha_k^2} + \frac{\|\bar{\eta}\|^2}{\alpha_k^2} + \frac{\|\bar{\gamma}\|^2}{\alpha_k^2}\right) \\ &= f(\hat{x}) + \rho\left(n - e^T \hat{y}\right) + \frac{1}{2}\left(\frac{\|\bar{\lambda}\|^2}{\alpha_k} + \frac{\|\bar{\mu}\|^2}{\alpha_k} + \frac{\|\bar{\eta}\|^2}{\alpha_k} + \frac{\|\bar{\gamma}\|^2}{\alpha_k}\right) \end{aligned}$$

From (4.33) we then obtain

$$f(u^k) + \rho\left(n - e^T w^k\right) \leq f(u^k) + \rho\left(n - e^T w^k\right) + \alpha_k \pi((u^k, w^k), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}, \alpha_k)$$

$$
\begin{aligned}
&= L((u^k, w^k), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) \\
&\leq L((\hat{x}, \hat{y}), \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) \\
&\leq f(\hat{x}) + \rho\left(n - e^T \hat{y}\right) + \frac{1}{2}\left(\frac{\|\bar{\lambda}\|^2}{\alpha_k} + \frac{\|\bar{\mu}\|^2}{\alpha_k} + \frac{\|\bar{\eta}\|^2}{\alpha_k} + \frac{\|\bar{\gamma}\|^2}{\alpha_k}\right).
\end{aligned}
$$

Letting $k \longrightarrow \infty$ yields $f(u) + \rho(n - e^T w) \leq f(\hat{x}) + \rho(n - e^T \hat{y})$. To arrive at the desired contradiction, it remains to prove that (u, w) is feasible for (4.30). Recall that $\|(u, w) - (\hat{x}, \hat{y})\| = \frac{\epsilon}{2}$. Thus, we only need to show that $(u, w) \in Z$. Suppose not. Then as measure of infeasibility we have $\|g(u)_+\|^2 + \|h(u)\|^2 + \|(w - e)_+\|^2 + \|u \circ w\|^2 > 0$. Since

$$
\left\{ \begin{array}{l} \left\|\left(g(u^k) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(u^k) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(w^k - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|u^k \circ w^k + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right\} \longrightarrow \|g(u)_+\|^2 + \|h(u)\|^2 + \|(w - e)_+\|^2 + \|u \circ w\|^2 > 0
$$

and

$$
\left\{ \begin{array}{l} \left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(\hat{x}) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|\hat{x} \circ \hat{y} + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right\} \longrightarrow \|g(\hat{x})_+\|^2 + \|h(\hat{x})\|^2 + \|(\hat{y} - e)_+\|^2 + \|\hat{x} \circ \hat{y}\|^2 = 0
$$

we can assume w.l.o.g. that there exists $d > 0$ such that for each $k \in \mathbb{N}$ we have

$$
\left(\begin{array}{l} \left\|\left(g(u^k) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(u^k) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(w^k - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|u^k \circ w^k + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right) > \left(\begin{array}{l} \left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(\hat{x}) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|\hat{x} \circ \hat{y} + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right) + d.
$$

Thus,

$$
\begin{aligned}
L((u^k, w^k), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) &= f(u^k) + \rho(n - e^T w^k) + \frac{\alpha_k}{2} \left(\begin{array}{l} \left\|\left(g(u^k) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(u^k) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(w^k - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|u^k \circ w^k + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right) \\
&> f(\hat{x}) + \rho(n - e^T \hat{y}) + \frac{\alpha_k}{2} \left(\begin{array}{l} \left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_k}\right)_+\right\|^2 + \left\|h(\hat{x}) + \frac{\bar{\mu}}{\alpha_k}\right\|^2 \\ + \left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_k}\right)_+\right\|^2 + \left\|\hat{x} \circ \hat{y} + \frac{\bar{\gamma}}{\alpha_k}\right\|^2 \end{array} \right) + \frac{\alpha_k d}{2} \\
&\quad + f(u^k) + \rho(n - e^T w^k) - f(\hat{x}) - \rho(n - e^T \hat{y}) \\
&= L((\hat{x}, \hat{y}), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) + \left(\begin{array}{c} \frac{\alpha_k d}{2} + f(u^k) + \rho(n - e^T w^k) \\ -f(\hat{x}) - \rho(n - e^T \hat{y}) \end{array} \right).
\end{aligned}
$$

Observe that $\left\{ \frac{\alpha_k d}{2} + f(u^k) + \rho(n - e^T w^k) - f(\hat{x}) - \rho(n - e^T \hat{y}) \right\} \longrightarrow \infty$. Hence, we can assume w.l.o.g. that it is positive for each $k \in \mathbb{N}$. Consequently,

$$
L((u^k, w^k), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k) > L((\hat{x}, \hat{y}), \bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}; \alpha_k).
$$

Thus, (u^k, w^k) cannot be a global minimiser of (4.31). This leads to a contradiction. Therefore, $(u, w) \in Z$ and we arrive at the desired contradiction since $(u, w) \neq (\hat{x}, \hat{y})$ and $(\hat{x}, \hat{y})$ is the unique global minimiser of (4.30). Hence, there exists $\hat{\alpha} > 0$ such that for each $\alpha \geq \hat{\alpha}$ (4.32) holds. Now if we pick $\alpha_1 \geq \hat{\alpha}$ and we let for each $k \in \mathbb{N}$ $\bar{\lambda}^k \in [0, \lambda_{\max}]^m$, $\bar{\mu}^k \in [\mu_{\min}, \mu_{\max}]^p$, $\bar{\eta}^k \in [0, \eta_{\max}]^n$, $\bar{\gamma}^k \in [\gamma_{\min}, \gamma_{\max}]^n$, and (x^k, y^k) as the global minimiser of the corresponding (4.31), then, since (4.32) holds, (x^k, y^k) is a local minimiser of

$$
\min_{x, y \in \mathbb{R}^n} L((x, y), \bar{\lambda}^k, \bar{\mu}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k)
$$

as $B_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$ is an open neighbourhood of (x^k, y^k) in which it minimises the above problem. Thus, $\nabla_{(x,y)}L((x^k, y^k), \bar{\lambda}^k, \bar{\mu}^k, \bar{\eta}^k, \bar{\gamma}^k; \alpha_k) = 0$ and (4.18) is trivially satisfied for any $\{\epsilon_k\} \longrightarrow 0$.

It remains to show that $\{x^k\} \longrightarrow \hat{x}$. Since $\{(x^k, y^k)\} \subseteq \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$, by the compactness of $\bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$, $\{(x^k, y^k)\}$ has a convergent subsequence in $\bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y}))$. By passing to this subsequence we can assume w.l.o.g. that $\{(x^k, y^k)\}$ converges i.e. $\exists(\bar{x}, \bar{y}) \in \bar{B}_{\frac{\epsilon}{2}}((\hat{x}, \hat{y})) : \{(x^k, y^k)\} \longrightarrow (\bar{x}, \bar{y})$. Now if $\{\alpha_k\}$ is unbounded, then by (S_3), we can assume w.l.o.g. that $\{\alpha_k\} \longrightarrow \infty$. Using similar argument as before we can prove that $(\bar{x}, \bar{y}) = (\hat{x}, \hat{y})$. Hence, in particular $\{x^k\} \longrightarrow \hat{x}$. Now consider the case where $\{\alpha_k\}$ is bounded. We shall prove that $(\hat{x}, \hat{y}) = (\bar{x}, \bar{y})$. Recall that $(\hat{x}, \hat{y})$ is the unique global minimiser of (4.30). Hence, if we can show that $(\bar{x}, \bar{y})$ is also a global minimiser of (4.30), then the claim follows from the uniqueness. The sequences $\{\bar{\lambda}^k\}$, $\{\bar{\mu}^k\}$, $\{\bar{\eta}^k\}$, and $\{\bar{\gamma}^k\}$ are all bounded. Thus, by passing to an appropriate subsequence we can assume w.l.o.g. that they all converge, i.e. $\exists(\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma}) : \{(\bar{\lambda}^k, \bar{\mu}^k, \bar{\eta}^k, \bar{\gamma}^k)\} \longrightarrow (\bar{\lambda}, \bar{\mu}, \bar{\eta}, \bar{\gamma})$. Furthermore, by (S_3), the boundedness of $\{\alpha_k\}$ implies that there exists $K \in \mathbb{N}$ such that $\alpha_k = \alpha_K$ for each $k \geq K$. Now by (4.19) we then have that

- $h(\bar{x}) = \lim_{k\longrightarrow\infty} h(x^k) = 0$,
- $\forall i \in \{1, \dots, m\} : \min\left\{-g_i(\bar{x}), \frac{\bar{\lambda}_i}{\alpha_K}\right\} = \lim_{k\longrightarrow\infty} U_i^k = 0 \implies g_i(\bar{x}) \leq 0$,
- $\forall i \in \{1, \dots, n\} : \min\left\{-(\bar{y}_i - 1), \frac{\bar{\eta}_i}{\alpha_K}\right\} = \lim k \longrightarrow \infty W_i^k = 0 \implies \bar{y}_i \leq 1$,
- $\bar{x}\circ\bar{y} = \lim_{k\longrightarrow\infty} x^k\circ y^k = 0$.

Hence, $(\bar{x}, \bar{y}) \in Z$ and is therefore feasible for (4.30). Observe that

- $\|\bar{x} - \hat{x}\| \leq \|(\bar{x}, \bar{y}) - (\hat{x}, \hat{y})\| \leq \frac{\epsilon}{2} < \epsilon \implies g_i(\bar{x}) < 0 \ \forall i \notin I_g(\hat{x})$,
- $\|\bar{y} - \hat{y}\| \leq \|(\bar{x}, \bar{y}) - (\hat{x}, \hat{y})\| \leq \frac{\epsilon}{2} < \epsilon \implies \bar{y}_i < 1 \ \forall i \in I_\pm(\hat{x})$.

The preceeding discussion then implies that

- $\forall i \notin I_g(\hat{x}) : \min\left\{-g_i(\bar{x}), \frac{\bar{\lambda}_i}{\alpha_K}\right\} = 0 \overset{-g_i(\bar{x})>0}{\implies} \frac{\bar{\lambda}_i}{\alpha_K} = 0$,
- $\forall i \in I_\pm(\hat{x}) : \min\left\{-(\bar{y}_i - 1), \frac{\bar{\eta}_i}{\alpha_K}\right\} = 0 \overset{1-\bar{y}_i>0}{\implies} \frac{\bar{\eta}_i}{\alpha_K} = 0$.

Now by definition we have for each $k \in \mathbb{N}$ that

$$f(x^k) + \rho(n - e^T y^k) + \frac{\alpha_k}{2}\begin{pmatrix} \left\|\left(g(x^k) + \frac{\bar{\lambda}^k}{\alpha_k}\right)_+\right\|^2 + \left\|h(x^k) + \frac{\bar{\mu}^k}{\alpha_k}\right\|^2 \\ + \left\|\left(y^k - e + \frac{\bar{\eta}^k}{\alpha_k}\right)_+\right\|^2 + \left\|x^k\circ y^k + \frac{\bar{\gamma}^k}{\alpha_k}\right\|^2 \end{pmatrix}$$

$$\leq f(\hat{x}) + \rho(n - e^T \hat{y}) + \frac{\alpha_k}{2}\begin{pmatrix} \left\|\left(g(\hat{x}) + \frac{\bar{\lambda}^k}{\alpha_k}\right)_+\right\|^2 + \left\|h(\hat{x}) + \frac{\bar{\mu}^k}{\alpha_k}\right\|^2 \\ + \left\|\left(\hat{y} - e + \frac{\bar{\eta}^k}{\alpha_k}\right)_+\right\|^2 + \left\|\hat{x}\circ\hat{y} + \frac{\bar{\gamma}^k}{\alpha_k}\right\|^2 \end{pmatrix}$$

Letting $k \longrightarrow \infty$ then yields

$$f(\bar{x}) + \rho(n - e^T \bar{y}) + \frac{\alpha_K}{2}\left(\left\|\left(g(\bar{x}) + \frac{\bar{\lambda}}{\alpha_K}\right)_+\right\|^2 + \left\|\frac{\bar{\mu}}{\alpha_K}\right\|^2 + \left\|\left(\bar{y} - e + \frac{\bar{\eta}}{\alpha_K}\right)_+\right\|^2 + \left\|\frac{\bar{\gamma}}{\alpha_K}\right\|^2\right)$$

$$\leq f(\hat{x}) + \rho(n - e^T \hat{y}) + \frac{\alpha_K}{2}\left(\left\|\left(g(\hat{x}) + \frac{\bar{\lambda}}{\alpha_K}\right)_+\right\|^2 + \left\|\frac{\bar{\mu}}{\alpha_K}\right\|^2 + \left\|\left(\hat{y} - e + \frac{\bar{\eta}}{\alpha_K}\right)_+\right\|^2 + \left\|\frac{\bar{\gamma}}{\alpha_K}\right\|^2\right)$$

which leads to

$$f(\bar{x})+\rho(n-e^T\bar{y})+\frac{\alpha_K}{2}\left(\left\|\left(g(\bar{x})+\frac{\bar{\lambda}}{\alpha_K}\right)_+\right\|^2+\left\|\left(\bar{y}-e+\frac{\bar{\eta}}{\alpha_K}\right)_+\right\|^2\right)$$
$$\leq f(\hat{x})+\rho(n-e^T\hat{y})+\frac{\alpha_K}{2}\left(\left\|\left(g(\hat{x})+\frac{\bar{\lambda}}{\alpha_K}\right)_+\right\|^2+\left\|\left(\hat{y}-e+\frac{\bar{\eta}}{\alpha_K}\right)_+\right\|^2\right)$$

Recall that

- $\forall i \notin I_g(\hat{x})$: $\frac{\bar{\lambda}_i}{\alpha_K}=0$. This implies that
  - $\max\left\{0, g_i(\hat{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right\}=\max\left\{0, g_i(\hat{x})\right\} \overset{g_i(\hat{x})<0}{=} 0$,
  - $\max\left\{0, g_i(\bar{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right\}=\max\left\{0, g_i(\bar{x})\right\} \overset{g_i(\bar{x})<0}{=} 0$.
- $\forall i \in I_\pm(\hat{x})$: $\frac{\bar{\eta}_i}{\alpha_K}=0$. This implies that
  - $\max\left\{0, \hat{y}_i-1+\frac{\bar{\eta}_i}{\alpha_K}\right\}=\max\left\{0, \hat{y}_i-1\right\} \overset{\hat{y}_i<1}{=} 0$,
  - $\max\left\{0, \bar{y}_i-1+\frac{\bar{\eta}_i}{\alpha_K}\right\}=\max\left\{0, \bar{y}_i-1\right\} \overset{\bar{y}_i<1}{=} 0$.

Hence,

$$f(\bar{x})+\rho(n-e^T\bar{y})+\frac{\alpha_K}{2}\left(\sum_{i\in I_g(\hat{x})}\left(g_i(\bar{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2+\sum_{i\in I_0(\hat{x})}\left(\bar{y}_i-1+\frac{\bar{\eta}_i}{\alpha_K}\right)_+^2\right)$$
$$\leq f(\hat{x})+\rho(n-e^T\hat{y})+\frac{\alpha_K}{2}\left(\sum_{i\in I_g(\hat{x})}\left(\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2+\sum_{i\in I_0(\hat{x})}\left(\frac{\bar{\eta}_i}{\alpha_K}\right)_+^2\right).$$

Now observe that since $(\bar{x},\bar{y})\in Z$

- $\forall i \in I_g(\hat{x}):\ \ g_i(\bar{x})\leq 0$. Thus,
  - if $g_i(\bar{x})=0$ then $\left(g_i(\bar{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2=\left(\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2$,
  - if $g_i(\bar{x})<0$ then $\min\left\{-g_i(\bar{x}), \frac{\bar{\lambda}_i}{\alpha_K}\right\}=0 \overset{-g_i(\bar{x})>0}{\Longrightarrow} \frac{\bar{\lambda}_i}{\alpha_K}=0$. Hence

$$\max\left\{0, g_i(\bar{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right\}=\max\{0, g_i(\bar{x})\} \overset{g_i(\bar{x})<0}{=} 0=\max\left\{0, \frac{\bar{\lambda}_i}{\alpha_K}\right\}.$$

  In both cases we have $\left(g_i(\bar{x})+\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2=\left(\frac{\bar{\lambda}_i}{\alpha_K}\right)_+^2$.

- $\forall i \in I_0(\hat{x}):\ \ \bar{y}_i\leq 1$. Thus,
  - if $\bar{y}_i=1$ then $\left(\bar{y}_i-1+\frac{\bar{\eta}_i}{\alpha_K}\right)_+^2=\left(\frac{\bar{\eta}_i}{\alpha_K}\right)_+^2$,
  - if $\bar{y}_i<1$ then $\min\left\{-(\bar{y}_i-1), \frac{\bar{\eta}_i}{\alpha_K}\right\}=0 \overset{1-\bar{y}_i>0}{\Longrightarrow} \frac{\bar{\eta}_i}{\alpha_K}=0$. Hence

$$\max\left\{0, \bar{y}_i-1+\frac{\bar{\eta}_i}{\alpha_K}\right\}=\max\{0, \bar{y}_i-1\} \overset{\bar{y}_i<1}{=} 0=\max\left\{0, \frac{\bar{\eta}_i}{\alpha_K}\right\}.$$

In both cases we have $\left(\bar{y}_i - 1 + \frac{\bar{\eta}_i}{\alpha_K}\right)_+^2 = \left(\frac{\bar{\eta}_i}{\alpha_K}\right)_+^2$.

This leads to

$$f(\bar{x}) + \rho(n - e^T\bar{y}) \le f(\hat{x}) + \rho(n - e^T\hat{y}).$$

Since $(\hat{x}, \hat{y})$ is the unique global minimiser of (4.30) and $(\bar{x}, \bar{y})$ is feasible for (4.30), then we must have $(\bar{x}, \bar{y}) = (\hat{x}, \hat{y})$. Thus, $\{x^k\} \longrightarrow \hat{x}$. This completes the proof. □

4.4.2 A Two-sided Scholtes Regularisation Method

To relax the orthogonality constraint $x \circ y = 0$ of (1.10) we shall now consider a variant of a two-sided regularisation method introduced by Scholtes, see [53], and further developed in [25]. Translating this method to the relaxed programme (1.10) means that instead of solving the problem directly, we solve a sequence of regularised problem NLP(T)

$$\begin{aligned} \min_{x,y}\ f(x) + \rho\left(n - e^T y\right) \quad \text{s.t.}\quad & g_i(x) \le 0 && \forall i = 1, \dots, m,\\ & h_i(x) = 0 && \forall i = 1, \dots, p,\\ & y_i \le 1 && \forall i = 1, \dots, n,\\ & x_i y_i \le t_i^+ && \forall i = 1, \dots, n,\\ & -x_i y_i \le t_i^- && \forall i = 1, \dots, n, \end{aligned} \tag{4.34}$$

with a vector of regularisation parameter $T = (t^+, t^-) \in (0, \infty)^{2n}$. Thus, instead of having only one regularisation parameter $t \in (0, \infty)$ going to zero, we introduce two parameters $t_i^+, t_i^- > 0$ for each pair (x_i, y_i). The idea behind this is to be able to drive only those bounds t_i^+, t_i^- to zero, which are needed to ensure feasibility of the limit of solutions of a sequence of NLP(T_k). More precisely, let $T_k > 0$ and (x^k, y^k) be a KKT-point of NLP(T_k) and $\sigma \in (0, 1)$. Then we update the regularisation parameter T_k as follows:

$$\left(t_i^{+,k+1}, t_i^{-,k+1}\right) = \begin{cases} \left(\sigma t_i^{+,k}, t_i^{-,k}\right) & \text{if } x_i^k y_i^k > 0,\\ \left(t_i^{+,k}, \sigma t_i^{-,k}\right) & \text{if } x_i^k y_i^k < 0,\\ \left(t_i^{+,k}, t_i^{-,k}\right) & \text{if } x_i^k y_i^k = 0. \end{cases} \tag{4.35}$$

The limit of such a sequence of KKT-points (x^k, y^k) is then feasible for (1.10). But compared to a one-sided relaxation method, the feasible set of NLP(T_k) might possess better properties as not all parameters t_i^+, t_i^- are necessarily driven to zero.

In the exact case, i.e. under the assumption that in each iteration $k \in \mathbb{N}$ we are able to compute an exact KKT-point (x^k, y^k) of NLP(T_k) we obtain the following convergence result.

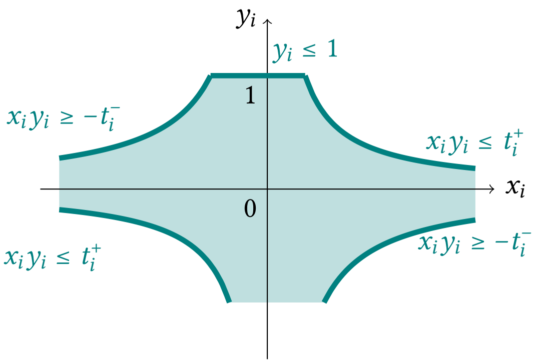


Figure 4.1: Illustration of the two-sided Scholtes regularisation method

Theorem 4.69. *Let $\{(x^k, y^k)\}$ be a sequence of KKT-points of NLP(T_k) where T_k is updated according to the rule* (4.35). *Suppose that $\{x^k\} \longrightarrow \hat{x}$. Then $\hat{x}$ is an SP-AKKT point of* (1.2).

We omit the proof since it is similar to the inexact case which we will handle next.

Theorem 4.70. *Let $\{\epsilon_k\} \downarrow 0$ and $\{(x^k, y^k)\}$ be a sequence of ϵ_k-stationary points of NLP(T_k) where T_k is updated according to the rule* (4.35). *Suppose that $\{x^k\} \longrightarrow \hat{x}$. Then $\hat{x}$ is an SP-AKKT point of* (1.2).

Proof. By assumption, for each $k \in \mathbb{N}$ there exists $(\lambda^k, \mu^k, \eta^k, \gamma^{+,k}, \gamma^{-,k}) \in \mathbb{R}^m \times \mathbb{R}^p \times (\mathbb{R}^n)^3$ such that

$$(sc_1)\ \left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^{+,k} y_i^k e_i - \sum_{i=1}^n \gamma_i^{-,k} y_i^k e_i \right\| \le \epsilon_k,$$

$$(sc_2)\ \left\| -\rho e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \gamma_i^{+,k} x_i^k e_i - \sum_{i=1}^n \gamma_i^{-,k} x_i^k e_i \right\| \le \epsilon_k,$$

$$(sc_3)\ \lambda_i^k \ge -\epsilon_k, \quad g_i(x^k) \le \epsilon_k, \quad |\lambda_i^k g_i(x^k)| \le \epsilon_k \quad \forall i = 1, \dots, m,$$

$$(sc_4)\ |h_i(x^k)| \le \epsilon_k \quad \forall = 1, \dots, p,$$

$$(sc_5)\ \eta_i^k \ge -\epsilon_k, \quad y_i^k - 1 \le \epsilon_k, \quad |\eta_i^k (y_i^k - 1)| \le \epsilon_k \quad \forall i = 1, \dots, n,$$

$$(sc_6)\ \gamma_i^{+,k} \ge -\epsilon_k, \quad x_i^k y_i^k - t_i^{+,k} \le \epsilon_k, \quad |\gamma_i^{+,k}(x_i^k y_i^k - t_i^{+,k})| \le \epsilon_k \quad \forall i = 1, \dots, n,$$

$$(sc_7)\ \gamma_i^{-,k} \ge -\epsilon_k, \quad -x_i^k y_i^k - t_i^{-,k} \le \epsilon_k, \quad |\gamma_i^{-,k}(-x_i^k y_i^k - t_i^{-,k})| \le \epsilon_k \quad \forall i = 1, \dots, n.$$

Observe that since $\{\epsilon_k\} \downarrow 0$, by letting $k \longrightarrow \infty$ we obtain from (sc_3) and (sc_4) that

$$g_i(\hat{x}) \le 0\ \forall i = 1, \dots, m \quad \wedge \quad h_i(\hat{x}) = 0\ \forall i = 1, \dots, p.$$

Hence, $\hat{x}$ is feasible for (1.2). Define

$$\begin{aligned} A^k &:= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^{+,k} y_i^k e_i - \sum_{i=1}^n \gamma_i^{-,k} y_i^k e_i, \\ B^k &:= -\rho e + \sum_{i=1}^n \eta_i^k e_i + \sum_{i=1}^n \gamma_i^{+,k} x_i^k e_i - \sum_{i=1}^n \gamma_i^{-,k} x_i^k e_i. \end{aligned} \tag{4.36}$$

By (sc_1) and (sc_2) we have $\{A^k\} \longrightarrow 0$ and $\{B^k\} \longrightarrow 0$. Now let $i \notin I_g(\hat{x})$. Then $g_i(\hat{x}) < 0$. Since $\{g_i(x^k)\} \longrightarrow g_i(\hat{x})$ we can assume w.l.o.g. that $g_i(x^k) < 0$ for each $k \in \mathbb{N}$. This implies that $|g_i(x^k)| > 0$ and thus, from (sc_3) we obtain for each $k \in \mathbb{N}$

$$0 \le |\lambda_i^k| \le \frac{\epsilon_k}{|g_i(x^k)|} \quad \Rightarrow \quad \{\lambda_i^k\} \longrightarrow 0 \quad \Rightarrow \{\lambda_i^k \nabla g_i(x^k)\} \longrightarrow 0.$$

Let us now define for each $k \in \mathbb{N}$ $\hat{\lambda}^k \in \mathbb{R}^m$ such that

$$\hat{\lambda}_i^k := \begin{cases} \lambda_i^k + \epsilon_k & \text{if } i \in I_g(\hat{x}), \\ 0 & \text{if } i \notin I_g(\hat{x}). \end{cases}$$

By (sc_3) $\{\hat{\lambda}_i^k\} \subseteq \mathbb{R}_+^m$. Now since $\{\epsilon_k\} \downarrow 0$, we then have for each $i \in I_g(\hat{x})$ that $\{\epsilon_k \nabla g_i(x^k)\} \longrightarrow 0$. From (4.36) we obtain

$$\bar{A}^k := A^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_g(\hat{x})} \epsilon_k \nabla g_i(x^k)$$

$$= \nabla f(x^k) + \sum_{i\in I_g(\hat{x})} (\lambda_i^k + \epsilon_k)\nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^{+,k} y_i^k e_i - \sum_{i=1}^{n} \gamma_i^{-,k} y_i^k e_i$$

$$= \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \gamma_i^{+,k} y_i^k e_i - \sum_{i=1}^{n} \gamma_i^{-,k} y_i^k e_i. \tag{4.37}$$

Observe that by definition we have $\{\bar{A}^k\} \longrightarrow 0$. Now let $i \in I_\pm(\hat{x})$. Since $\hat{x}_i \neq 0$ and $\{x_i^k\} \longrightarrow \hat{x}_i$, we can assume w.l.o.g. that $x_i^k \neq 0$ for each $k \in \mathbb{N}$. From (sc_6) and (sc_7) we obtain that $-t_i^{-,k} - \epsilon_k \leq x_i^k y_i^k \leq t_i^{+,k} + \epsilon_k$ for each $k \in \mathbb{N}$. Now since $\sigma \in (0,1)$ by (4.35) $\{t_i^{+,k}\}$ and $\{t_i^{-,k}\}$ are nonincreasing sequences. Hence, in particular, we have for each $k \in \mathbb{N}$ that $t_i^{+,k} \leq t_i^{+,1}$ and $-t_i^{-,1} \leq -t_i^{-,k}$. Thus we have $-t_i^{-,1} - \epsilon_k \leq x_i^k y_i^k \leq t_i^{+,1} + \epsilon_k$ and therefore, the convergence of $\{\epsilon_k\}$ immediately implies the boundedness of $\{x_i^k y_i^k\}$. By passing to a subsequence, we can assume w.l.o.g. that $\{x_i^k y_i^k\}$ converges, i.e. $\exists z \in \mathbb{R} : \{x_i^k y_i^k\} \longrightarrow z$. We claim that $z = 0$. Suppose not. We only consider the case where $z > 0$. The case where $z < 0$ can be handled analogously. Now since $\{x_i^k y_i^k\} \longrightarrow z > 0$, we can assume w.l.o.g. that $x_i^k y_i^k > 0$ for each $k \in \mathbb{N}$. Since $\sigma \in (0,1)$, by (4.35) it follows that $\{t_i^{+,k}\} \longrightarrow 0$. By (sc_6) we have $x_i^k y_i^k \leq t_i^{+,k} + \epsilon_k$ for each $k \in \mathbb{N}$ and hence, by letting $k \longrightarrow \infty$ we obtain $z \leq 0$, a contradiction. Hence, $z = 0$. Then we have

$$\lim_{k\to\infty} y_i^k = \lim_{k\to\infty} \frac{1}{x_i^k} x_i^k y_i^k = \lim_{k\to\infty} \frac{1}{x_i^k} \cdot \lim_{k\to\infty} x_i^k y_i^k = \frac{1}{\hat{x}_i} \cdot 0 = 0.$$

Thus, $\{|y_i^k - 1|\} \longrightarrow 1$ and we can therefore assume w.l.o.g. $|y_i^k - 1| > 0$ that for each $k \in \mathbb{N}$. Hence, from (sc_5) we obtain

$$0 \leq |\eta_i^k| \leq \frac{\epsilon_k}{|y_i^k - 1|} \quad \Longrightarrow \quad \{\eta_i^k\} \longrightarrow 0.$$

From (4.36) we have

$$B_i^k + \rho - \eta_i^k = (\gamma_i^{+,k} - \gamma_i^{-,k}) x_i^k \quad \Longrightarrow \quad \frac{B_i^k + \rho - \eta_i^k}{x_i^k} \cdot y_i^k = (\gamma_i^{+,k} - \gamma_i^{-,k}) y_i^k.$$

Since $\{B_i^k\} \longrightarrow 0$, we have $\left\{ \frac{B_i^k + \rho - \eta_i^k}{x_i^k} \right\} \longrightarrow \frac{\rho}{\hat{x}_i}$. Thus,

$$\left\{ (\gamma_i^{+,k} - \gamma_i^{-,k}) y_i^k \right\} = \left\{ \frac{B_i^k + \rho - \eta_i^k}{x_i^k} \cdot y_i^k \right\} \longrightarrow \frac{\rho}{\hat{x}_i} \cdot 0 = 0.$$

Now define for each $k \in \mathbb{N}$ $\hat{\gamma}^k \in \mathbb{R}^n$ such that

$$\hat{\gamma}_i^k := \begin{cases} (\gamma_i^{+,k} - \gamma_i^{-,k}) y_i^k & \text{if } i \in I_0(\hat{x}), \\ 0 & \text{if } i \in I_\pm(\hat{x}). \end{cases}$$

From (4.37) we obtain

$$\begin{aligned} \hat{A}^k &:= \bar{A}^k - \sum_{i\in I_\pm(\hat{x})} (\gamma_i^{+,k} - \gamma_i^{-,k}) y_i^k e_i \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i\in I_0(\hat{x})} (\gamma_i^{+,k} - \gamma_i^{-,k}) y_i^k e_i \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^{n} \hat{\gamma}_i^k e_i. \end{aligned}$$

Since $\{\hat{A}\} \longrightarrow 0$, the assertion then follows. □

An immediate consequence of Theorem 4.38 is the following

Corollary 4.71. *If, in addition to the assumptions in Theorem 4.70, $\hat{x}$ also satisfies SP-AKKT-regularity, then $\hat{x}$ is an SP-KKT point of* (1.2).

In order to solve the regularised problems NLP(T), we need to make sure that they, possibly contrary to the original relaxed problem (1.10), satisfy suitable standard constraint qualifications. To simplify the notation we define for each feasible point (x, y) of NLP(T)

$$I_+((x,y)) := \{i \in \{1,\dots,n\} \mid x_i y_i = t_i^+\} \quad \wedge \quad I_-((x,y)) := \{i \in \{1,\dots,n\} \mid -x_i y_i = t_i^-\}.$$

Since $T > 0$ obviously we have

$$I_+((x,y)) \cap I_-((x,y)) = \emptyset. \tag{4.38}$$

Theorem 4.72. *Let $\hat{x} \in X$ such that SP-CPLD holds at $\hat{x}$. Then there exists $\epsilon > 0$ such that for all $T > 0$ the standard CPLD for NLP(T) is satisfied at every feasible point (x, y) of NLP(T) with $x \in B_\epsilon(\hat{x})$.*

Proof. Suppose not. Then there exist $\{T_k\} \subseteq (0,\infty)^{2n}$ and $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$ such that $\{x^k\} \longrightarrow \hat{x}$ and for each $k \in \mathbb{N}$ (x^k, y^k) is a feasible point of NLP(T_k) which does not satisfy CPLD. Then for each $k \in \mathbb{N}$ there exist $I_1^k \subseteq I_g(x^k)$, $I_2^k \subseteq \{1,\dots,p\}$, $I_3^k \subseteq I_1(y^k)$, $I_4^k \subseteq I_+((x^k,y^k))$, $I_5^k \subseteq I_-((x^k,y^k))$, and

$$\left(\lambda_i^k \geq 0 \ (i \in I_1^k),\ \mu_i^k \ (i \in I_2^k),\ \eta_i^k \geq 0 \ (i \in I_3^k),\ \gamma_i^{+,k} \geq 0 \ (i \in I_4^k),\ \gamma_i^{-,k} \geq 0 \ (i \in I_5^k)\right) \neq 0 \tag{4.39}$$

such that

$$0 = \sum_{i \in I_1^k} \lambda_i^k \begin{bmatrix} \nabla g_i(x^k) \\ 0 \end{bmatrix} + \sum_{i \in I_2^k} \mu_i^k \begin{bmatrix} \nabla h_i(x^k) \\ 0 \end{bmatrix} + \sum_{i \in I_3^k} \eta_i^k \begin{bmatrix} 0 \\ e_i \end{bmatrix} + \sum_{i \in I_4^k} \gamma_i^{+,k} \begin{bmatrix} y_i^k e_i \\ x_i^k e_i \end{bmatrix} - \sum_{i \in I_5^k} \gamma_i^{-,k} \begin{bmatrix} y_i^k e_i \\ x_i^k e_i \end{bmatrix}, \tag{4.40}$$

as well as $(u^k, w^k) \in B_{\frac{1}{k}}((x^k, y^k))$ such that the family

$$\left\{ \begin{matrix} \begin{bmatrix} \nabla g_i(u^k) \\ 0 \end{bmatrix} (i \in I_1^k), & \begin{bmatrix} \nabla h_i(u^k) \\ 0 \end{bmatrix} (i \in I_2^k), & \begin{bmatrix} 0 \\ e_i \end{bmatrix} (i \in I_3^k), \\ \begin{bmatrix} w_i^k e_i \\ u_i^k e_i \end{bmatrix} (i \in I_4^k), & \begin{bmatrix} -w_i^k e_i \\ -u_i^k e_i \end{bmatrix} (i \in I_5^k) & \end{matrix} \right\} \text{ is linearly independent.} \tag{4.41}$$

Now observe that the index sets are all finite. Hence, by passing to a subsequence we can assume w.l.o.g. that there exist $I_1 \subseteq \{1,\dots,m\}$, $I_2 \subseteq \{1,\dots,p\}$, $I_3, I_4, I_5 \subseteq \{1,\dots,n\}$ such that for each $k \in \mathbb{N}$ we have $I_j = I_j^k$ for each $j \in \{1,\dots,5\}$. Observe that by (4.38) we have $I_4^k \cap I_5^k = \emptyset$ for each $k \in \mathbb{N}$ and hence, we also have $I_4 \cap I_5 = \emptyset$. From (4.40) we then obtain

$$0 = \sum_{i \in I_1} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_2} \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_4} \gamma_i^{+,k} y_i^k e_i - \sum_{i \in I_5} \gamma_i^{-,k} y_i^k e_i \tag{4.42}$$

and

$$0 = \sum_{i \in I_3} \eta_i^k e_i + \sum_{i \in I_4} \gamma_i^{+,k} x_i^k e_i - \sum_{i \in I_5} \gamma_i^{-,k} x_i^k e_i. \tag{4.43}$$

Let $i \in I_3$. We now claim that $\eta_i^k = 0$ for each $k \in \mathbb{N}$. Suppose not. Then there exists $k \in \mathbb{N}$ such that $\eta_i^k > 0$. We differentiate between two cases.
<u>Case 1</u>: $i \notin I_4 \dot{\cup} I_5$
From (4.43) we then obtain $\eta_i^k = 0$, a contradiction.
<u>Case 2</u>: $i \in I_4 \dot{\cup} I_5$
Then we have $i \in I_4 \ \dot{\vee}\ i \in I_5$. Assume that $i \in I_4$. The other case can be dealt with analogously. Since $I_3 \subseteq I_1(y^k)$, we have $y_i^k = 1$. Furthermore, since $I_4 \subseteq I_+((x^k, y^k))$, we also have $x_i^k y_i^k = t_i^{+,k}$. Hence, $x_i^k = t_i^{+,k} > 0$. From (4.43), since $\gamma_i^{+,k} \geq 0$ we obtain

$$0 < \eta_i^k = -\gamma_i^{+,k} x_i^k \leq 0.$$

This leads to a contradiction.
Hence, we conclude that $\eta_i^k = 0$ for each $k \in \mathbb{N}$. Consequently, from (4.43) we then obtain

$$\gamma_i^{+,k} x_i^k = 0 \; \forall i \in I_4 \quad \wedge \quad \gamma_i^{-,k} x_i^k = 0 \; \forall i \in I_5$$

since $I_4 \cap I_5 = \emptyset$. Now let $i \in I_4$. Since $I_4 \subseteq I_+((x^k, y^k))$ we then have $x_i^k y_i^k = t_i^{+,k} > 0$ and hence, in particular, $x_i^k \neq 0$. This implies that $\gamma_i^{+,k} = 0$ for each $k \in \mathbb{N}$. Similarly we obtain for each $i \in I_5$ that $\gamma_i^{-,k} = 0 \; \forall k \in \mathbb{N}$. Thus, from (4.42) we have for each $k \in \mathbb{N}$ that

$$0 = \sum_{i \in I_1} \lambda_i^k \nabla g_i(x^k) + \sum_{i \in I_2} \mu_i^k \nabla h_i(x^k). \tag{4.44}$$

By (4.39) we then have $(\lambda_i^k \; (i \in I_1), \; \mu_i^k \; (i \in I_2)) \neq 0$. The sequence $\left\{ \frac{(\lambda_i^k \; (i \in I_1), \; \mu_i^k \; (i \in I_2))}{\|(\lambda_i^k \; (i \in I_1), \; \mu_i^k \; (i \in I_2))\|} \right\}$ is then well defined and bounded. By passing to a subsequence, we can assume w.l.o.g. that it converges, i.e.

$$\exists (\hat{\lambda}_i \; (i \in I_1), \; \hat{\mu}_i \; (i \in I_2)) \neq 0 \; : \; \left\{ \frac{(\lambda_i^k \; (i \in I_1), \; \mu_i^k \; (i \in I_2))}{\|(\lambda_i^k \; (i \in I_1), \; \mu_i^k \; (i \in I_2))\|} \right\} \longrightarrow (\hat{\lambda}_i \; (i \in I_1), \; \hat{\mu}_i \; (i \in I_2)).$$

Observe that since $\lambda_i^k \geq 0$, we then have $\hat{\lambda}_i \geq 0$ as well. Dividing (4.44) by the norm and letting $k \longrightarrow \infty$ then yields

$$0 = \sum_{i \in I_1} \hat{\lambda}_i \nabla g_i(\hat{x}) + \sum_{i \in I_2} \hat{\mu}_i \nabla h_i(\hat{x}).$$

Observe that for each $k \in \mathbb{N}$ we have $I_1 \subseteq I_g(x^k)$. Hence, for each $i \in I_1$ we have $g_i(x^k) = 0$ and therefore, by letting $k \longrightarrow \infty$, $g_i(\hat{x}) = 0$. Thus, $I_1 \subseteq I_g(\hat{x})$. Furthermore, by definition we have $\{u^k\} \longrightarrow \hat{x}$ and by (4.41)

$$\{\nabla g_i(u^k) \; (i \in I_1), \; \nabla h_i(u^k) \; (i \in I_2)\} \;\; \text{is linearly independent.}$$

This contradicts the assumption that SP-CPLD holds at $\hat{x}$. The proof is complete. □

4.5 Numerical Experiments

In this section we shall benchmark the performances of ALGENCAN, the adaptation of the original Scholtes regularisation method [53] to (1.10), and the two-sided Scholtes regularisation method from Section 4.4.2 against the global solver CPLEX [24]. The only difference between the original Scholtes regularisation method and the two-sided version from Section 4.4.2 is that for the original Scholtes method we decrease both regularisation parameters t^+ and t^- in each iteration. Hence, the proof of Theorem 4.70 can be easily adapted for the original Scholtes method.

For our numerical tests, we consider the following sparse robust portfolio optimisation problem which is adapted from [18, Section 4]

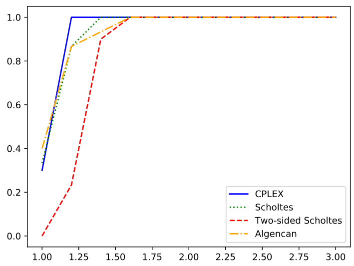
$$\min_{x \in \mathbb{R}^n} \; c_\beta \sqrt{x^T Q x} - \mu^T x + \rho \|x\|_0 \quad \text{s.t.} \quad e^T x = 1, \; 0 \leq x \leq u, \tag{4.45}$$

where Q and μ are the covariance matrix as well as the mean of n possible assets, $e^T x = 1$ is the budget constraint, and c_β is as given in [18, Table 1]. The only difference between (4.45) and the problem considered in [18] is that here we instead penalise the l_0-norm with a sparsity parameter ρ. Just like in [18], the test problems are generated using the data from [32] and consist of 3 different dimensions, namely $n = 200, 300$, and 400. In our experiments we set $\beta = 0.9$ and $\rho = 1.0$ and we considered VaR, CVaR, RVaR, and RCVaR. These result in 360 test problems. All experiments were conducted using Python together with the Numpy and the Scipy libraries. The constant c_β is computed in the case of VaR and CVaR using Scipy. We applied ALGENCAN, the Scholtes method, and the two-sided Scholtes variant to the relaxed reformulation (1.10) of (4.45). In order to solve (4.45) with CPLEX, we employed the following mixed-integer quadratically constrained reformulation of (4.45)

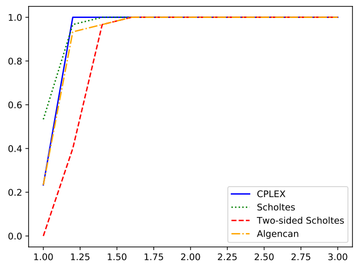
$$\min_{x,z,w,v} \; c_\beta v - \mu^T x + \rho e^T z \quad \text{s.t.} \quad \begin{aligned} & e^T x = 1, \\ & 0 \leq x \leq u \circ z, \\ & z \in \{0,1\}^n, \\ & 0 \leq v, \\ & w = Q^{\frac{1}{2}} x, \\ & v^2 \geq w^T w. \end{aligned} \tag{4.46}$$

For each test problem we called CPLEX through the DocPlex interface, ran it with a time limit of 300 seconds and set `cplex_parameters.emphasis.mip` to 1. As a start vector, following [18] we used $x^0 = 0$ and $z^0 = e$. For each test problem we observed that CPLEX always hit the time limit. As a sub-solver for both regularisation methods we used the for academic use freely available ESA SQP solver WORHP version 1.14 [22] called through its Python interface. We called WORHP with user-supplied sparse gradients of the objective functions, sparse Jacobian of the constraints, as well as the sparse Hessian of the Lagrangian. Moreover, we set the parameter `MoreRelax` to `True`. Since $e^T z$ corresponds to $n - e^T y$ in (1.10), as a start vector for both regularisation methods we used $x^0 = 0$ and $y^0 = 0$. We terminated both regularisation methods if either $T_k < 10^{-8}$ or $\|x^k \circ y^k\|_\infty \leq 10^{-6}$ and σ was chosen to be 0.1. As for ALGENCAN, we used ALGENCAN 2.4.0 compiled with MA57 library [37] and called through its Python interface with user-supplied gradients of the objective functions, sparse Jacobian of the constraints, as well as sparse Hessian of the Lagrangian. Throughout the experiments we skipped the acceleration steps as these would have turned ALGENCAN into a Newton method instead of Algorithm 4.59 in the vicinity of a solution. In our preliminary tests, using $x^0 = 0$ as the initial value for x, ALGENCAN kept generating VEVALHL warnings in the beginning even though it eventually managed to converge. Indeed, if $x = 0$ we would then have to deal with division by 0 when evaluating the Jacobian and the Hessian of the objective function. It seems that ALGENCAN, in contrast to WORHP, had a hard time coping with such degenerate cases. Thus, for ALGENCAN, as a start vector we instead used x^0 = `numpy.ones(n) / numpy.nan_to_num(numpy.inf)`, which roughly corresponds to $5.562684646268003 \cdot 10^{-309} \cdot e$, and $y^0 = 0$. Additionally, we converted any NaN and $\pm\infty$ encountered by ALGENCAN into a finite value using `numpy.nan_to_num`. For each test problem we also used the keywords `TRUST-REGIONS-INNER-SOLVER` and `HESSIAN-APPROXIMATION-IN-CG` in the ALGENCAN's parameter specification file. This, if the author has understood correctly, means that we set the trust-region method [10] as the default inner solver and the truncated Newton method with Hessian approximation as the fallback solver if ALGENCAN encountered difficulties in computing the true Hessian. For the most part this leads to a good result for ALGENCAN. The only exception is problem *orl400_05_b* for RVaR. Hence, for this problem only we replaced `HESSIAN-APPROXIMATION-IN-CG` with `INCREMENTAL-QUOTIENTS-IN-CG` in the parameter specification file which lead to a massive improvement in terms of the attained objective function value. In our experiments we also observed that ALGENCAN converged the fastest as one would expect.

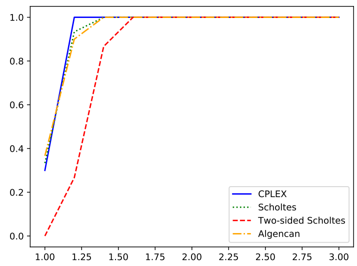
Let us now present the results of our experiments. Just like in Section 3.5, as a performance measure for the considered methods we compared the attained objective function values and generated a performance profile as suggested in [27], where we set the objective function value of a method for a problem to be ∞ if the method failed to find a feasible point of the problem within a tolerance of 10^{-6}. As can be seen from Figure 4.2, for $n = 200$, the two-sided Scholtes regularisation method seems to be the least successful solver and ALGENCAN came as the second worst. The Scholtes regularisation method seems to be competitive with CPLEX and even managed to clearly outperform CPLEX for RCVaR200. Let us now consider the case where $n \in \{300, 400\}$. Here the two-sided Scholtes regularisation method behaved for the most part in a similar way to its original counterpart with the two-sided variant having the edge over the original in the cases of RCVaR300 and RCVaR400. These two methods outperformed both ALGENCAN and CPLEX. As for ALGENCAN, with the exceptions of VaR300 and RCVaR300, it managed to outperform CPLEX.



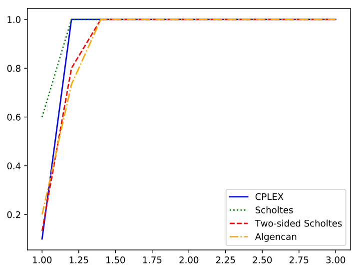
(a) VaR and $n = 200$



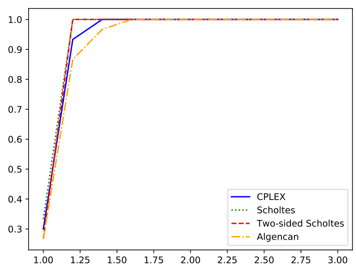
(b) CVaR and $n = 200$



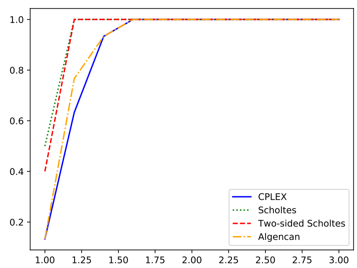
(c) RVaR and $n = 200$



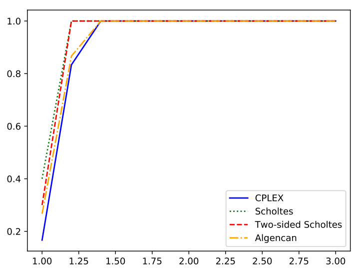
(d) RCVaR and $n = 200$



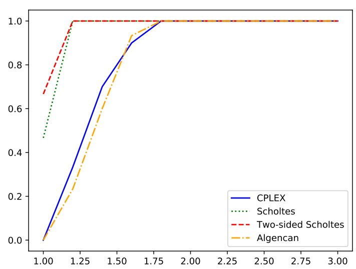
(e) VaR and $n = 300$



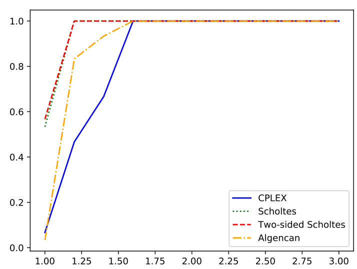
(f) CVaR and $n = 300$



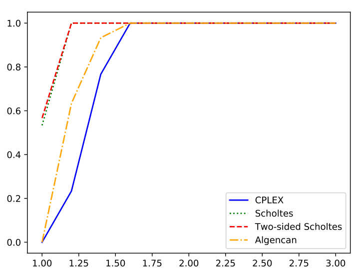
(g) RVaR and $n = 300$



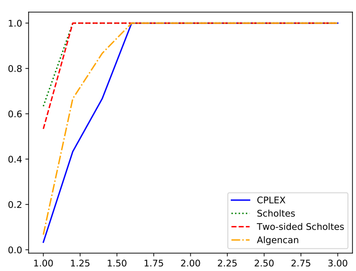
(h) RCVaR and $n = 300$



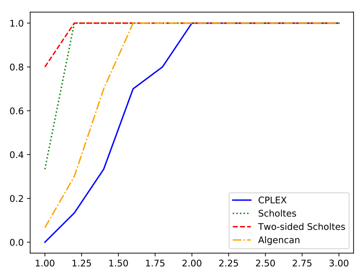
(i) VaR and $n = 400$



(j) CVaR and $n = 400$



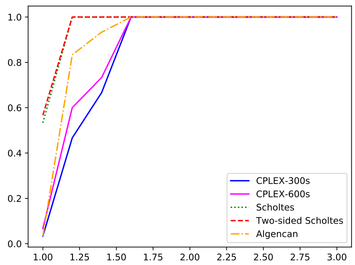
(k) RVaR and $n = 400$



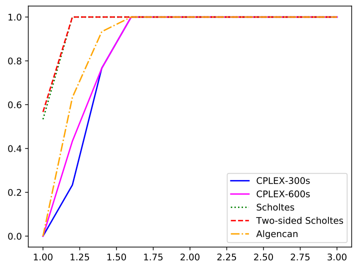
(l) RCVaR and $n = 400$

Figure 4.2: Comparing the performance of ALGENCAN, the regularisation methods, and CPLEX for (4.45)

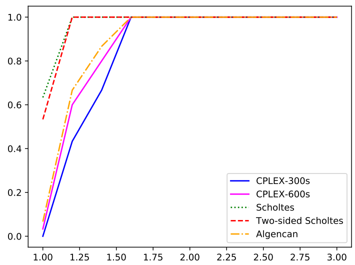
Since CPLEX did not perform well for $n = 400$ within the given 300 seconds time limit, we tried to increase the time limit to 600 seconds for this case to see if there is any improvement. However, even after doubling the time limit, we observed that CPLEX seemed to mostly get stuck and could not improve the attained objective function values. Consequently, as can be seen from Figure 4.3, CPLEX still gets outperformed by the other solvers.



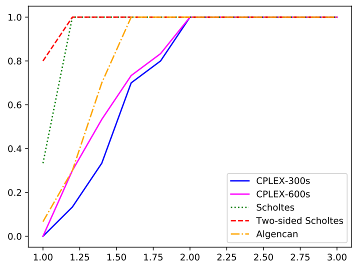
(a) VaR and $n = 400$



(b) CVaR and $n = 400$



(c) RVaR and $n = 400$



(d) RCVaR and $n = 400$

Figure 4.3: Performance plot of the objective function values for $n = 400$ with a time limit of 600 seconds for CPLEX

For our numerical tests, in terms of the objective function value, we can conclude that the original Scholtes regularisation method has the best overall performance. This is even more remarkable given that it always managed to converge to a good solution within a small fraction of the time limit that is allocated for CPLEX.

Chapter 5

Final Remarks

Let us now discuss some possible future research directions for (1.1) and (1.2). In Chapter 3 we have derived sequential optimality conditions for (1.1) and investigated some of the methods which can theoretically converge towards a point satisfying these conditions when applied to the relaxed reformulation (1.9). As highlighted in [21], however, (1.9) has a major drawback, namely that a local minimiser of (1.9) may not be a local minimiser of (1.1). Thus, it seems natural to ask whether we can construct an algorithm which can be applied directly to (1.1) and produces a point satisfying at least one of the sequential optimality conditions introduced in Chapter 3. The proof of Theorem 3.9 provides us with an algorithmic insight, namely that we may be able to combine techniques from [11, 15] to construct such an algorithm. In each iteration we only penalise the nonlinear constraints of (1.1). The objective function of the resulting subproblem is continuously differentiable and $\mathcal{S}$ is the only constraint left. Thus, the subproblem can theoretically then be handled using methods from [11]. It is then easy to see that if we are able to compute a BF-vector of the subproblem in each iteration and the thus generated sequence has a feasible limit point, this limit point must at least be a CC-AM-stationary point. The question now is whether the subproblem is computationally tractable. We will leave this for future research.

In Chapter 4 we have seen that in some cases like Example 4.8, (1.2) may possess too many local minima. Thus, it seems natural to ask if we can derive a necessary global optimality condition for (1.2). The relaxed reformulation (1.10) may offer us such possibility. In Chapter 4 we have seen that the solution sets of (1.2) and (1.10) coincide. Now observe that the constraints associated with the auxiliary variable y in (1.10) are all polynomial in nature. Thus, if the constraints g_i and h_i are polynomial as well, then techniques from [39] could perhaps be applied to derive global optimality conditions for (1.2) which hopefully can then be incorporated as a stopping criterium for solvers like ALGENCAN and WORHP to prevent convergence to a suboptimal solution. We shall leave this as future work.

Appendix A

Equivalence of Sequential Optimality Conditions

Here we shall take a look at a sequential optimality condition called *AW-stationarity* which was introduced in [44]. This condition was derived utilising the relaxed reformulation of (1.1) from [21], i.e.

$$\min_{x,y} f(x) \quad \text{s.t.} \quad x \in X, \;\; n - e^T y \leq s, \;\; 0 \leq y \leq e, \;\; x \circ y = 0, \tag{A.1}$$

which is simply (1.9) along with the constraint $y \geq 0$. Let us now compare CC-AM-stationarity with AW-stationarity. First we recall the definition of AW-stationarity from [44].

Definition A.1. *Let* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ *be feasible for* (A.1)*. We say that* $(\hat{x}, \hat{y})$ *is approximately weakly stationary (AW-stationary) for* (A.1) *iff there exist sequences* $\{(x^k, y^k)\} \subseteq \mathbb{R}^n \times \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, $\{\zeta_k\} \subseteq \mathbb{R}_+$, $\{\gamma^k\} \subseteq \mathbb{R}^n$, $\{\nu^k\} \subseteq \mathbb{R}^n$, *and* $\{\eta^k\} \subseteq \mathbb{R}^n_+$ *such that*

(a) $\left\{(x^k, y^k)\right\} \longrightarrow (\hat{x}, \hat{y})$,

(b) $\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} \longrightarrow 0$,

(c) $\left\{-\zeta_k e - \sum_{i=1}^n \nu_i^k e_i + \sum_{i=1}^n \eta_i^k e_i\right\} \longrightarrow 0$,

(d) $\forall i \in \{1, \dots, m\} : \left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0$,

(e) $\left\{\min\{-(n - s - e^T y^k), \zeta_k\}\right\} \longrightarrow 0$,

(f) $\forall i \in \{1, \dots, n\} : \left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0$,

(g) $\forall i \in \{1, \dots, n\} : \left\{\min\{y_i^k, |\nu_i^k|\}\right\} \longrightarrow 0$,

(h) $\forall i \in \{1, \dots, n\} : \left\{\min\{-(y_i^k - 1), \eta_i^k\}\right\} \longrightarrow 0$.

Next we derive an equivalent formulation of CC-AM-stationarity.

Proposition A.2. *Let* $\hat{x} \in \mathbb{R}^n$ *be feasible for* (1.1)*. Then* $\hat{x}$ *is CC-AM-stationary iff there exist sequences* $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, *and* $\{\gamma^k\} \subseteq \mathbb{R}^n$ *such that*

(a) $\{x^k\} \longrightarrow \hat{x}$,

(b) $\left\{\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i\right\} \longrightarrow 0$,

(c) $\forall i \in \{1, \dots, m\}$: $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0$,

(d) $\forall i \in \{1, \dots, n\}$: $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0$.

Proof. "$\Longrightarrow$": Assume first that $\hat{x}$ is CC-AM-stationary. We only need to prove that the corresponding sequences also satisfy conditions (c) and (d) in Proposition A.2. Let $i \notin I_g(\hat{x})$. We then have

$$g_i(\hat{x}) < 0 \quad \text{and} \quad \lambda_i^k = 0 \; \forall k \in \mathbb{N}.$$

Now since $\{g_i(x^k)\} \longrightarrow g_i(\hat{x})$, we can assume w.l.o.g. that $g_i(x^k) < 0 \; \forall k \in \mathbb{N}$. Hence,

$$\min\{\underbrace{-g_i(x^k)}_{>0}, \underbrace{\lambda_i^k}_{=0}\} = \lambda_i^k = 0 \; \forall k \in \mathbb{N},$$

which immediately implies that

$$\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0.$$

Now let $i \in I_g(\hat{x})$. Then $g_i(\hat{x}) = 0$. By assumption we have $\lambda_i^k \geq 0 \; \forall k \in \mathbb{N}$. We now claim that $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0$. Let $\epsilon > 0$. Since $\{g_i(x^k)\} \longrightarrow g_i(\hat{x}) = 0$, there exists $K \in \mathbb{N}$ such that $\forall k \geq K$ we have $|g_i(x^k)| < \epsilon$. Now let $k \geq K$. Suppose that $\min\{-g_i(x^k), \lambda_i^k\} = -g_i(x^k)$. Then we have

$$|\min\{-g_i(x^k), \lambda_i^k\}| = |-g_i(x^k)| = |g_i(x^k)| < \epsilon.$$

If on the other hand $\min\{-g_i(x^k), \lambda_i^k\} = \lambda_i^k$, then since

$$0 \leq \lambda_i^k \leq -g_i(x^k) \leq |g_i(x^k)| < \epsilon,$$

we have that

$$|\min\{-g_i(x^k), \lambda_i^k\}| = |\lambda_i^k| = \lambda_i^k < \epsilon.$$

In both cases we obtain that

$$|\min\{-g_i(x^k), \lambda_i^k\} - 0| < \epsilon.$$

Hence, we conclude that $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0$. Now let $i \in I_{\pm}(\hat{x})$. Then we have

$$|\hat{x}_i| > 0 \quad \text{and} \quad \gamma_i^k = 0 \; \forall k \in \mathbb{N}.$$

Since $\{x_i^k\} \longrightarrow \hat{x}_i$, we can then assume w.l.o.g. that $|x_i^k| > 0 \; \forall k \in \mathbb{N}$. Thus,

$$\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} = \{|\gamma_i^k|\} = \{0\} \longrightarrow 0.$$

Suppose now that $i \in I_0(\hat{x})$. We claim that $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0$. Let $\epsilon > 0$. Then since $\{x_i^k\} \longrightarrow 0$ there exists $K \in \mathbb{N}$ such that $\forall k \geq K$ we have $|x_i^k| < \epsilon$. This then immediately implies that

$$|\min\{|x_i^k|, |\gamma_i^k|\}| = \min\{|x_i^k|, |\gamma_i^k|\} \leq |x_i^k| < \epsilon.$$

Hence, we have $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0$.

"$\Longleftarrow$": Suppose now that there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that conditions (a) - (d) in Proposition A.2 hold. Define for each $k \in \mathbb{N}$

$$A^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i.$$

Then $\{A^k\} \longrightarrow 0$. Now let $i \notin I_g(\hat{x})$. Since $\{-g_i(x^k)\} \longrightarrow -g_i(\hat{x}) > 0$ and $\left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0$ we can assume w.l.o.g. that $\forall k \in \mathbb{N}$

$$\min\{-g_i(x^k), \lambda_i^k\} < -g_i(x^k) \implies \min\{-g_i(x^k), \lambda_i^k\} = \lambda_i^k$$

which implies that
$$\left\{\lambda_i^k\right\} = \left\{\min\{-g_i(x^k), \lambda_i^k\}\right\} \longrightarrow 0.$$

Define now for each $k \in \mathbb{N}$ $\hat{\lambda}^k \in \mathbb{R}^m$ such that

$$\hat{\lambda}_i^k := \begin{cases} 0 & \text{if } i \notin I_g(\hat{x}), \\ \lambda_i^k & \text{if } i \in I_g(\hat{x}). \end{cases}$$

Since $\{\lambda^k\} \subseteq \mathbb{R}_+^m$, then clearly we also have $\{\hat{\lambda}^k\} \subseteq \mathbb{R}_+^m$. By definition we then have $\forall i \notin I_g(\hat{x})$ that $\hat{\lambda}_i^k = 0$ $\forall k \in \mathbb{N}$. Next we define

$$\begin{aligned} B^k &:= A^k - \sum_{i \notin I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) \\ &= \nabla f(x^k) + \sum_{i \in I_g(\hat{x})} \lambda_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i \\ &= \nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \gamma_i^k e_i. \end{aligned}$$

Now since by the preceeding discussion we have $\{\lambda_i^k\} \longrightarrow 0$ $\forall i \notin I_g(\hat{x})$, we then have $\{\lambda_i^k \nabla g_i(x^k)\} \longrightarrow 0 \cdot \nabla g_i(\hat{x}) = 0$ $\forall i \notin I_g(\hat{x})$ and hence, $\{B^k\} \longrightarrow 0$. Next let $i \in I_\pm(\hat{x})$. Then since $\{|x_i^k|\} \longrightarrow |\hat{x}_i| > 0$ and $\left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0$ we can assume w.l.o.g. that $\forall k \in \mathbb{N}$

$$\min\{|x_i^k|, |\gamma_i^k|\} < |x_i^k| \implies \min\{|x_i^k|, |\gamma_i^k|\} = |\gamma_i^k|.$$

Thus,
$$\left\{|\gamma_i^k|\right\} = \left\{\min\{|x_i^k|, |\gamma_i^k|\}\right\} \longrightarrow 0,$$
which implies that $\{\gamma_i^k\} \longrightarrow 0$. Define now $\forall k \in \mathbb{N}$ $\hat{\gamma}^k \in \mathbb{R}^n$ such that

$$\hat{\gamma}_i^k := \begin{cases} 0 & \text{if } i \in I_\pm(\hat{x}), \\ \gamma_i^k & \text{if } i \in I_0(\hat{x}). \end{cases}$$

Then clearly we have $\forall i \in I_\pm(\hat{x})$ that $\hat{\gamma}_i^k = 0$ $\forall k \in \mathbb{N}$. Now define for each $k \in \mathbb{N}$

$$\begin{aligned} C^k &:= B^k - \sum_{i \in I_\pm(\hat{x})} \gamma_i^k e_i \\ &= \nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i \in I_0(\hat{x})} \gamma_i^k e_i \\ &= \nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla g_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla h_i(x^k) + \sum_{i=1}^n \hat{\gamma}_i^k e_i \end{aligned}$$

Observe that since by the preceeding discussion we have $\{\gamma_i^k\} \longrightarrow 0$ $\forall i \in I_\pm(\hat{x})$, we then have $\{\gamma_i^k e_i\} \longrightarrow 0$ $\forall i \in I_\pm(\hat{x})$ and hence, $\{C^k\} \longrightarrow 0$. Thus, we conclude that $\hat{x}$ is CC-AM-stationary with the corresponding sequences $\{x^k\}$, $\{\hat{\lambda}^k\}$, $\{\mu^k\}$, and $\{\hat{\gamma}^k\}$. □

Recall from [21] that if $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ is feasible for (A.1), then $\hat{x}$ is feasible for (1.1). An immediate consequence of Definition A.1 and Proposition A.2 is then the following.

Theorem A.3. *Let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ be a feasible point of* (A.1). *If $(\hat{x}, \hat{y})$ is AW-stationary, then $\hat{x}$ is CC-AM-stationary.*

For the converse we obtain the following.

Theorem A.4. *Let $\hat{x} \in \mathbb{R}^n$ be a feasible point of* (1.1). *If $\hat{x}$ is a CC-AM-stationary point, then for any $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for* (A.1) *it follows that $(\hat{x}, \hat{y})$ is AW-stationary.*

Proof. Assume that $\hat{x}$ is CC-AM-stationary. Then there exist sequences $\{x^k\} \subseteq \mathbb{R}^n$, $\{\lambda^k\} \subseteq \mathbb{R}^m_+$, $\{\mu^k\} \subseteq \mathbb{R}^p$, and $\{\gamma^k\} \subseteq \mathbb{R}^n$ such that conditions (a) - (d) in Proposition A.2 hold. Now let $\hat{y} \in \mathbb{R}^n$ such that $(\hat{x}, \hat{y})$ is feasible for (A.1). Then we can simply define for each $k \in \mathbb{N}$

$$y^k := \hat{y}, \quad \zeta_k := 0, \quad \nu^k := 0, \quad \eta^k := 0.$$

Conditions (a) - (d) and (f) in Definition A.1 are trivially satisfied. Now since $(\hat{x}, \hat{y})$ is feasible for (A.1), we then have $0 \leq -(n - e^T \hat{y} - s) = -(n - e^T y^k - s)$ $\forall k \in \mathbb{N}$. Hence,

$$\min\{\underbrace{-(n - e^T y^k - s)}_{\geq 0}, \underbrace{\zeta_k}_{=0}\} = \zeta_k = 0 \ \forall k \in \mathbb{N}$$

and therefore, $\left\{\min\{-(n - e^T y^k - s), \zeta_k\}\right\} \longrightarrow 0$. Now since $(\hat{x}, \hat{y})$ is feasible for (A.1), we also have for each $i \in \{1, \dots, n\}$ that $0 \leq \hat{y}_i = y_i^k$ $\forall k \in \mathbb{N}$. Hence,

$$\min\{\underbrace{y_i^k}_{\geq 0}, \underbrace{|\nu_i^k|}_{=0}\} = |\nu_i^k| = 0 \ \forall k \in \mathbb{N}$$

and therefore, $\left\{\min\{y_i^k, |\nu_i^k|\}\right\} \longrightarrow 0$. Moreover, the feasibility of $(\hat{x}, \hat{y})$ also implies that for each $i \in \{1, \dots, n\}$ we have $0 \leq 1 - \hat{y}_i = 1 - y_i^k$ $\forall k \in \mathbb{N}$. Hence,

$$\min\{\underbrace{1 - y_i^k}_{\geq 0}, \underbrace{\eta_i^k}_{=0}\} = \eta_i^k = 0$$

and therefore, $\left\{\min\{1 - y_i^k, \eta_i^k\}\right\} \longrightarrow 0$. This completes the proof. □

An obvious advantage of CC-AM-stationarity over AW-stationarity is that it does not depend on the artificial variable y. Indeed, as we have already shown, CC-AM-stationarity is a genuine optimality condition for the original problem (1.1).

Bibliography

[1] J. M. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 21–36, New York, 1967. Wiley.

[2] H. Amann and J. Escher. *Analysis 1.* Grundstudium Mathematik. Birkhäuser, Basel, 3rd edition, 2006.

[3] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125:473–483, 2005.

[4] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18:1286–1309, 2007.

[5] 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.

[6] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60:627–641, 2011.

[7] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26:96–110, 2016.

[8] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43:693–717, 2018.

[9] 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.

[10] M. Andretta, E. G. Birgin, and J. M. Martínez. Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization. *Optimization*, 54(3):305–325, 2005.

[11] A. Beck and Y. C. Eldar. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. *SIAM Journal on Optimization*, 23(3):1480–1509, 2013.

[12] D. P. Bertsekas and A. E. Ozdaglar. Pseudonormality and a Lagrange multiplier theory for constrained optimization. *Journal of Optimization Theory and Applications*, 114(2):287–343, 2002.

[13] D. Bertsimas and R. Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43:1–22, 2009.

[14] D. Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74:121–140, 1996.

[15] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization.* SIAM, Philadelphia, 2014.

[16] E. G. Birgin, C. A. Floudas, and J. M. Martínez. Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Mathematical Programming*, 125:139–162, 2010.

[17] E. G. Birgin, D. Fernández, and J. M. Martínez. The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. *Optimization Methods and Software*, 27(6):1001–1024, 2012.

[18] M. Branda, M. Bucher, M. Červinka, and A. 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*, 70:503–530, 2018.

[19] M. Bucher. *Optimality Conditions and Numerical Methods for a Continuous Reformulation of Cardinality Constrained Optimization Problems.* PhD thesis, Technische Universität Darmstadt, 2018.

[20] O. P. Burdakov, C. Kanzow, and A. Schwartz. On Sparse Optimization Problems: Theory and Algorithm. Unpublished, 2014.

[21] O. P. Burdakov, C. Kanzow, and A. Schwartz. Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26:397–425, 2016.

[22] C. Büskens and D. Wassel. The ESA NLP solver WORHP. In *Modeling and optimization in space engineering*, volume 73 of *Springer Optim. Appl.*, pages 85–110. Springer, New York, 2013.

[23] M. Červinka, C. Kanzow, and A. Schwartz. Constraint qualifications and optimality conditions of cardinality-constrained optimization problems. *Mathematical Programming*, 160(1):353–377, 2016.

[24] IBM ILOG CPLEX. *V12.10: User's Manual for CPLEX.* International Business Machines Corporation.

[25] 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.

[26] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen, and M. Sciandrone. A concave optimization based approach for sparse portfolio selection. *Optimization Methods and Software*, 27:983–1000, 2012.

[27] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2, Ser. A):201–213, 2002.

[28] H. Dong, M. Ahn, and J.-S. Pang. Structural properties of affine sparsity constraints. *Mathematical Programming*, 176:95–135, 2019.

[29] M. Durea and R. Strugariu. *An Introduction to Nonlinear Optimization Theory.* De Gruyter Open, Warsaw, 2014.

[30] M. Feng, J. E. Mitchell, J.-S. Pang, X. Shen, and A. Wächter. Complementarity formulations of l_0-norm optimization problems. *Pacific Journal of Optimization*, 14(2):273–305, 2018.

[31] D. Fernández and M. V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. *SIAM Journal on Optimization*, 22(2):384–407, 2012.

[32] A. Frangione and C. Gentile. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Operations Research Letters*, 35:181–185, 2007.

[33] C. Geiger and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben.* Springer, Berlin, Heidelberg, 2002.

[34] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

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

[36] T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137:257–288, 2013.

[37] HSL. *A collection of Fortran codes for large scale scientific computation.* `http://www.hsl.rl.ac.uk/`.

[38] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM Journal on Optimization*, 22(4):1579–1606, 2012.

[39] V. Jeyakumar and G. Y. Li. Necessary global optimality conditions for nonlinear programming problems with polynomial constraints. *Mathematical Programming*, 126:393–399, 2011.

[40] 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.

[41] C. Kanzow and A. Schwartz. The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementary constraints revisited. *Mathematics of Operations Research*, 40(21):253–275, 2015.

[42] C. Kanzow, A. B. Raharja, and A. Schwartz. An augmented Lagrangian method for cardinality-constrained optimization problems. Technical report, Institute of Mathematics, University of Würzburg, 2020.

[43] C. Kanzow, A. B. Raharja, and A. Schwartz. Sequential optimality conditions for cardinality-constrained optimization problems with applications. Technical report, Institute of Mathematics, University of Würzburg, 2021.

[44] E. H. M. Krulikovski, A. A. Ribeiro, and M. Sachine. A sequential optimality condition for mathematical programs with cardinality constraints. *ArXiv e-prints*, 2020. ArXiv:2008.03158.

[45] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen, and M. Sciandrone. A concave optimization-based approach for sparse portfolio selection. *Optimization Methods and Software*, 27:983–1000, 2012.

[46] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17: 37–47, 1967.

[47] P. Mehlitz. Stationarity conditions and constraint qualifications for mathematical programs with switching constraints. *Mathematical Programming*, pages 1–38, 2019.

[48] W. Murray and H. Shek. A local relaxation method for the cardinality constrained portfolio optimization problem. *Computational Optimization and Applications*, 53:681–709, 2012.

[49] L.-L. Pan, N.-H. Xiu, and S.-L. Zhou. On solutions of sparsity constrained optimization. *Journal of the Operations Research Society of China*, 3:421–439, 2015.

[50] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, 10:963–981, 2000.

[51] A. Ramos. Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications, and algorithmic consequences. *Optimization Methods and Software*, 2019. doi: 10.1080/10556788.2019.1702661.

[52] R. T. Rockafellar and R. J-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1st edition, 1998.

[53] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

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

[55] X. Sun, X. Zheng, and D. Li. Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. *Journal of the Operations Research Society of China*, 1:55–77, 2013.

[56] X. Zheng, X. Sun, D. Li, and J. Sun. Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear dc approach. *Computational Optimization and Applications*, 59:379–397, 2014.