On the best achievable quality of limit points of augmented Lagrangian schemes*

R. Andreani† G. Haeser‡ L.M. Mito‡ A. Ramos§ L.D. Secchin¶

July 23, 2020 (updated September 26, 2021)

Abstract

The optimization literature is vast in papers dealing with improvements on the global convergence of augmented Lagrangian schemes. Usually, the results are based on weak constraint qualifications, or, more recently, on *sequential* optimality conditions obtained via penalization techniques. In this paper we propose a somewhat different approach, in the sense that the algorithm itself is used in order to formulate a new optimality condition satisfied by its feasible limit points. With this tool at hand, we present several new properties and insights on limit points of augmented Lagrangian schemes, in particular, characterizing the strongest possible global convergence result for the safeguarded augmented Lagrangian method.

Keywords: Nonlinear optimization, Augmented Lagrangian methods, Optimality conditions.
AMS subject classifications: 90C46, 90C30, 65K05

1 Introduction

In this paper we deal with the following problem:

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } h(x) = 0,\ g(x) \leqslant 0, \end{aligned} \tag{NLP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions.

There is a variety of ways in which the *quality* of a general scheme for solving (NLP) may be measured. Most often, numerical stability, execution time, convergence rate, or other measures would be employed. However, in this paper, we are interested in the quality of the limit points of the sequences generated by the scheme, in the sense of how close they are to being necessarily optimal. This is an important complementary analysis of the reliability of an algorithm. Here we restrict our analysis and conclusions only to augmented Lagrangian schemes. Usually, in most descriptions of algorithms for (NLP), the question of the quality of its limit points is answered in a simple way: a constraint qualification (CQ) such as the linear independence CQ or the Mangasarian-Fromovitz CQ is assumed at all feasible points of (NLP) and it is shown that all feasible limit points of a sequence generated by the algorithm satisfies the Karush-Kuhn-Tucker (KKT) conditions. Several weaker CQs have been considered in the recent years yielding stronger global convergence results based on weaker constraint qualifications such as RCPLD [10], CPG [11], CCP [15], Quasinormality [4], among others.

*This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPES (grant 116/2019), FAPESP (grants 2018/24293-0, 2017/18308-2 and 2017/17840-2), CNPq (grants 301888/2017-5, 303427/2018-3, 404656/2018-8, 438185/2018-8 and 307270/2019-0) and PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016).

†Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil. E-mail: andreani@unicamp.br

‡Department of Applied Mathematics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil. E-mail: ghaeser@ime.usp.br, leokoto@ime.usp.br

§Department of Mathematics, Federal University of Paraná, 81531-980, Curitiba, PR, Brazil. Email: albertoramos@ufpr.br

¶Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, ES, Brazil. E-mail: leonardo.secchin@ufes.br

One may restate a global convergence result that a KKT point is achieved under a given CQ by saying that feasible limit points of an algorithm satisfy “KKT or not-CQ”, meaning either KKT holds or that particular CQ is violated. This has the advantage that the latter is a true necessary optimality condition, satisfied at all local minimizers of (NLP), while KKT on its own may fail for specific problems. This simple approach has lead to the definition of different but genuine necessary optimality conditions that imply “KKT or not-CQ” but that are more tailored to global convergence analysis of algorithms. In particular, these have been called *sequential optimality conditions* [4, 7, 17, 40] and the sequences needed for checking the validity of the condition are precisely the primal and dual sequences generated by the algorithm.

The simplest example of a sequential optimality condition is the so-called Approximate-KKT (AKKT) condition, which is said to hold at a feasible point x^* whenever one may find a primal sequence $\{x^k\} \subset \mathbb{R}^n$, $x^k \to x^*$ and a dual sequence $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}^p_+$ such that

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j:g_j(x^*)=0} \mu_j^k \nabla g_j(x^k) \to 0.$$

The simple observation that most reasonable algorithms generate sequences with this property is enough to recover most global convergence results to a KKT point under a constraint qualification. However, this is a poor analysis of the quality of x^*, as the sequences generated by the algorithm have a much more specific form. For a concrete example, consider the problem of minimizing $(x_1 - 1)^2 + (x_2 - 1)^2$ subject to $x_1 \geqslant 0, x_2 \geqslant 0, x_1 x_2 \leqslant 0$. Here, the only solutions are $(1, 0)$ and $(0, 1)$ but all feasible points satisfy AKKT [4]. However, it is known that an augmented Lagrangian method can only accumulate around $(1, 0), (0, 1)$, or $(0, 0)$ [32]. Also, even though an AKKT sequence is always generated, different augmented Lagrangian schemes will have different convergence properties (see, e.g., [36]) and in this paper we will investigate the impact on the quality of the limit point in view of several variants of augmented Lagrangian schemes, in particular taking into account how the dual sequence is computed.

In some sense, there is a mismatch between the sequences generated by the augmented Lagrangian algorithm and the sequences proved to exist converging to a local minimizer, since the latter is frequently built using a pure external penalty method on a regularized problem, and not the augmented Lagrangian. In this paper we will use *the sequence itself*, generated by the augmented Lagrangian algorithm, to attest optimality of a limit point. More precisely, we show that *all* local minimizers are limit points of a sequence generated by the algorithm. This is a somewhat different approach from previous global convergence results via sequential optimality conditions, in the sense that there is no need to employ an optimality condition disconnected from the algorithm. The algorithm itself is used to study the quality of its limit points. With this tool at hand, we are able to investigate further what are the important features of the algorithm, in the sense that this property may be lost when modifying it. Surprisingly, we show that the algorithm may fail to achieve some local minimizers if exact stationary points are found in each subproblem or exact feasible points, hence, it is paramount to allow some degree of error when solving the subproblems. We also show that some simplified variants of the augmented Lagrangian method are equivalent to the standard algorithm in terms of the quality of its limit points, while a variant including a penalty parameter for each constraint induces a weaker optimality condition.

1.1 Contributions of this article

In this paper, we introduce a mathematical description of all feasible limit points of an augmented Lagrangian algorithm that also takes into account the sequences it generates (Definition 1). Then, using this description:

- we prove that being a feasible limit point of the algorithm is a necessary condition for local optimality (Theorem 2);
- we characterize the weakest constraint qualification required for proving global convergence to KKT points (Section 4);
- we prove that a growth control over the penalty parameter does not affect the quality of its limit points (Section 3.1), even though it provides a computational gain of stability;

- we provide strong evidences that suggest that the safeguarded augmented Lagrangian method is equivalent to the classical external penalty method in terms of the quality of their feasible limit points (Section 3.2 and Corollary 2);
- we show that using a different penalty parameter for each constraint is theoretically worse than using a common parameter (Section 3.4);
- we discuss the theoretical implications of employing a safeguarding technique for updating Lagrange multipliers (Theorem 5);
- we show that forcing exact feasibility or stationarity over the subproblems may possibly cause undesired effects on its convergence;
- we discuss other implementation details and their effects on the quality of limit points.

1.2 Notation

Our work environment is $\mathbb{R}^n$ equipped with the Euclidean norm, which is defined as $\|x\|_2 \doteq \sqrt{x_1^2 + \cdots + x_n^2}$ for every $x \in \mathbb{R}^n$. To attest algorithmic convergence, we employ the norm $\|x\|_\infty \doteq \max\{|x_i|\colon i \in \{1,\ldots,n\}\}$. The nonnegative orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}^n_+ \doteq \{x \in \mathbb{R}^n\colon x_i \geqslant 0,\ \forall i \in \{1,\ldots,n\}\}$ and, similarly, its positive orthant is $\mathbb{R}^n_{++} \doteq \{x \in \mathbb{R}^n\colon x_i > 0,\ \forall i \in \{1,\ldots,n\}\}$. It is well-known that the projection of a point $x \in \mathbb{R}^n$ onto $\mathbb{R}^n_+$ is given by $[x]_+ \doteq (\max\{0,x_1\},\ldots,\max\{0,x_n\})$, and that $\nabla\|[x]_+\|_2^2 = 2[x]_+$, for every x.

The *Lagrangian function* of (NLP) is defined as $L(x,\lambda,\mu) \doteq f(x) + h(x)^T\lambda + g(x)^T\mu$, where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$. Also, its gradient with respect to x is given by $\nabla_x L(x,\lambda,\mu) \doteq \nabla f(x) + \nabla h(x)^T\lambda + \nabla g(x)^T\mu$.

We denote sequences labeled by a variable x and indexed by a set $J \subseteq \mathbb{N}$ by $\{x^k\}_{k\in J}$, and in this context, $x^k \rightsquigarrow x^j$ means that x^j is the successor of x^k, where $j,k \in J$.

2 The PHR augmented Lagrangian algorithm

There exist many augmented Lagrangian variants in the literature, but the main subject of our analyses is the one known as the *Powell-Hestenes-Rockafellar* (PHR) algorithm [30, 43, 47], which is characterized by the following shifted penalty function:

$$L_{\rho,\bar\lambda,\bar\mu}(x) \doteq f(x) + \frac{\rho}{2}\left[\left\|\frac{\bar\lambda}{\rho} + h(x)\right\|_2^2 + \left\|\left[\frac{\bar\mu}{\rho} + g(x)\right]_+\right\|_2^2\right], \tag{1}$$

where $\rho > 0$ and $\bar\mu \geqslant 0$. It is worth mentioning that a practical comparison among 65 distinct augmented Lagrangian variants was presented in [20], and the PHR was observed to have the best performance in their experiments.

As usual for penalty-type methods, the core idea behind the PHR algorithm is to minimize $L_{\rho,\bar\lambda,\bar\mu}(x)$ successively until some stopping criterion is satisfied, increasing ρ whenever needed, and updating $\bar\lambda$ and $\bar\mu$ in a suitable way. There are many possible implementations for it, but in this paper our analyses revolve around a consolidated implementation known as Algencan [2, 22], which has a robust practical implementation provided by the TANGO project (`www.ime.usp.br/~egbirgin/tango`) and a good track record of applications (for a description of applications in several fields, we refer to [22]).

The algorithm is usually presented as in Algorithm 1.

Remark 1. *Strictly speaking,* Algencan *allows for additional box-constraints in its subproblems, which are solved by an active-set strategy. However, this is not relevant to our analysis, so we abuse the notation by naming Algorithm 1 as* Algencan*.*

In many practical situations, $\mathcal{B}$ is defined as a box in the form $[\lambda_{\min},\lambda_{\max}]^m \times [0,\mu_{\max}]^p$, where $\lambda_{\min} \leqslant \lambda_{\max}$ and $\mu_{\max} \geqslant 0$ are given. Also, the usual choice of $(\bar\lambda^{k+1},\bar\mu^{k+1})$ is the projection of $(\lambda^k \doteq \bar\lambda^k + \rho_k h(x^k), \mu^k \doteq [\bar\mu^k + \rho_k g(x^k)]_+)$ onto $\mathcal{B}$, but a priori it can be any other element

Algorithm 1 ALGENCAN

Input: Parameters $\tau \in (0,1)$ and $\gamma > 1$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial penalty parameter $\rho_1 > 0$, a compact set $\mathcal{B} \subseteq \mathbb{R}^m \times \mathbb{R}^p_+$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\|\nabla L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x^k)\|_\infty \leqslant \varepsilon_k.$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \left(h(x^k), \min\left\{ -g(x^k), \frac{\bar{\mu}^k}{\rho_k} \right\} \right).$$

Then,

a. If $k = 1$ or $\|V^k\|_\infty \leqslant \tau \|V^{k-1}\|_\infty$, set $\rho_{k+1} \doteq \rho_k$;

b. Otherwise, take ρ_{k+1} such that $\rho_{k+1} \geqslant \gamma\rho_k$.

Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k+1$ and go to Step 1.

---

of $\mathcal{B}$. This technique is often called *safeguarding* and, in this context, $\bar{\lambda}^k$ and $\bar{\mu}^k$ are called *safeguarded multipliers*. Some advantages of safeguarding are discussed in Section 3.3. The vector V^k defined by (3) is a joint measure of feasibility and complementarity that is meant to control the growth of the penalty parameter through *Step 2*. In particular, note that ALGENCAN even allows $\{\rho_k\}_{k\in\mathbb{N}}$ to be bounded, which is a specially meaningfull feature when (NLP) is convex [46]. As a matter of fact, there is no universal rule for updating ρ_k, but it is important to keep in mind that, from the numerical point of view, the difficulty of solving the subproblems grows when ρ_k increases. Also, in situations where the augmented Lagrangian method is performing well, the penalty parameter can even be allowed to decrease [21], mainly because moderate values of ρ_k usually mean a better behavior of the box-constraint solver for the subproblem. Another work that deals with the possibility of decreasing the penalty parameter is [25]. This topic is out of the scope of this paper, but nevertheless we choose to give some degree of freedom for ρ_k for the sake of generality.

2.1 The sequences generated by the augmented Lagrangian method

In order to build a rigorous analysis of ALGENCAN, it is fundamental to introduce a consistent mathematical characterization of its output sequences and their limit points. We get inspiration from the structure of sequential optimality conditions to encapsulate every feasible outcome of ALGENCAN in the following definition:

Definition 1. *We say that x^* is an* Augmented Lagrangian AKKT *(AL-AKKT) point if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that*

$$\nabla_x L(x^k, \lambda^k, \mu^k) \to 0 \quad \text{and} \quad V^k \to 0 \tag{2}$$

where $\lambda^k \doteq \bar{\lambda}^k + \rho_k h(x^k)$, $\mu^k \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$, and

$$V^k \doteq \left(h(x^k), \min\left\{ -g(x^k), \frac{\bar{\mu}^k}{\rho_k} \right\} \right). \tag{3}$$

In this context, we say that $\{x^k\}_{k\in\mathbb{N}}$ is a (primal) AL-AKKT sequence associated with the dual sequence $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$.

Although it is intuitively clear that Definition 1 is indeed a complete characterization of the output sequences of ALGENCAN, this is not obvious whatsoever. To make this relationship clear, we define an *instance* of ALGENCAN as a two-phase procedure consisting of:

1. Choosing the initial parameters: $\tau, \gamma, \{\varepsilon^k\}_{k\in\mathbb{N}}, \rho_1, \mathcal{B}$, and $(\bar{\lambda}_1, \bar{\mu}_1)$;
2. Running ALGENCAN, producing sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\lambda^k, \mu^k\}_{k\in\mathbb{N}}$.

Also, we say that an instance of ALGENCAN *generates* a given point x^* when x^* is an accumulation point of its output sequence $\{x^k\}_{k\in\mathbb{N}}$.

In the following lines, we present a formal argument that establishes the correspondence between the algorithmic representation of ALGENCAN and Definition 1.

Theorem 1. *For each AL-AKKT point x^*, there is an instance of* ALGENCAN *such that its output sequence $\{x^k\}_{k\in\mathbb{N}}$ has x^* as an accumulation point. Conversely, every feasible accumulation point generated by* ALGENCAN *is AL-AKKT.*

Proof. Let $\{x^k\}_{k\in\mathbb{N}}$ be an AL-AKKT sequence convergent to x^*, with associated safeguarded multipliers $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$, and parameters $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$. We will split the proof in two cases: the first one is when $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, which means that for any $\gamma > 1$, there is a subsequence of it such that $\rho_{k+1} \geqslant \gamma\rho_k$ for every k. We denote the index set of this subsequence by $K = \{i_1, i_2, i_3, \dots\} \subset \mathbb{N}$. Then, fix $\tau \in (0,1)$ and define the index set

$$K_= \doteq \{i_k \geqslant 2 \mid \|V^{i_k}\|_\infty \leqslant \tau\|V^{i_{k-1}}\|_\infty\} \subset K.$$

This set may be finite or infinite. Despite of this, we define a set I and new sequences $\{\tilde{x}^j\}_{j\in I}$, $\{\tilde{\lambda}^j\}_{j\in I}$, $\{\tilde{\mu}^j\}_{j\in I}$, $\{\tilde{\rho}_j\}_{j\in I}$ in the following way:

1. Start with $\tilde{x}^{i_1} \doteq x^{i_1}$, $\tilde{\lambda}^{i_1} \doteq \bar{\lambda}^{i_1}$, $\tilde{\mu}^{i_1} \doteq \bar{\mu}^{i_1}$ and $\tilde{\rho}_{i_1} \doteq \rho_{i_1}$;
2. For each $i_k \in K\backslash\{i_1\}$,
   (a) if $i_k \in K_=$ then define the next elements $\tilde{x}^{i_k-\frac{1}{2}} \doteq x^{i_k}$, $\tilde{\lambda}^{i_k-\frac{1}{2}} \doteq \bar{\lambda}^{i_k}$, $\tilde{\mu}^{i_k-\frac{1}{2}} \doteq \bar{\mu}^{i_k}$ and $\tilde{\rho}_{i_k-\frac{1}{2}} \doteq \rho_{i_k}$;
   (b) independently if i_k is in $K_=$ or not, define the next elements $\tilde{x}^{i_k} \doteq x^{i_k}$, $\tilde{\lambda}^{i_k} \doteq \bar{\lambda}^{i_k}$, $\tilde{\mu}^{i_k} \doteq \bar{\mu}^{i_k}$ and $\tilde{\rho}_{i_k} \doteq \rho_{i_k}$,

and I is defined as K with duplicated elements whenever they belong to $K_=$. For example, if $K_= = \{i_2, i_3, i_6\}$ then

$$\{\tilde{x}^j\}_I = \{\ x^{i_1},\ x^{i_2-\frac{1}{2}} = x^{i_2},\ x^{i_2},\ x^{i_3-\frac{1}{2}} = x^{i_3},\ x^{i_3},\ x^{i_4},\ x^{i_5},\ x^{i_6-\frac{1}{2}} = x^{i_6},\ x^{i_6},\ \dots\}$$

and $I = \{i_1, i_2, i_2, i_3, i_3, i_4, i_5, i_6, i_6, \dots\}$ (for simplicity, we abuse the notation here by duplicating indexes). The other sequences are analogous. In other words, we duplicate an element of every sequence whenever its index is in $K_=$. In the next lines, the vectors $\tilde{V}^j$ are defined as in (3) for those new sequences.

There are the following possibilities for any two consecutive elements:

- $\tilde{x}^{i_{k-1}} \rightsquigarrow \tilde{x}^{i_k-\frac{1}{2}}$. In this case, $i_k \in K_=$ and then $\|\tilde{V}^{i_k-\frac{1}{2}}\|_\infty = \|\tilde{V}^{i_k}\|_\infty \leqslant \tau\|\tilde{V}^{i_{k-1}}\|_\infty$. Note that, by definition, $\tilde{\rho}_{i_k} = \tilde{\rho}_{i_k-\frac{1}{2}}$;
- $\tilde{x}^{i_k-\frac{1}{2}} \rightsquigarrow \tilde{x}^{i_k}$. Here, $\|\tilde{V}^{i_k}\|_\infty = \|\tilde{V}^{i_k-\frac{1}{2}}\|_\infty > \tau\|\tilde{V}^{i_k-\frac{1}{2}}\|_\infty$. The next element is $\tilde{x}^{i_{k+1}}$ or $\tilde{x}^{i_{k+1}-\frac{1}{2}}$. In both cases, $(\tilde{\rho}_{i_{k+1}-\frac{1}{2}} =)\tilde{\rho}_{i_{k+1}} > \tilde{\rho}_{i_k}$;
- $\tilde{x}^{i_{k-1}} \rightsquigarrow \tilde{x}^{i_k}$. In this case $i_k \notin K_=$ and thus $\|\tilde{V}^{i_k}\|_\infty > \tau\|\tilde{V}^{i_{k-1}}\|_\infty$. Again, the next element is $\tilde{x}^{i_{k+1}}$ or $\tilde{x}^{i_{k+1}-\frac{1}{2}}$, from which $(\tilde{\rho}_{i_{k+1}-\frac{1}{2}} =)\tilde{\rho}_{i_{k+1}} > \tilde{\rho}_{i_k}$.

Thus, we conclude that in this case the new sequences satisfy, consecutively, all requirements of ALGENCAN with safeguarded multipliers $\{\bar{\lambda}^j\}_{j\in I}$ and $\{\bar{\mu}^j\}_{j\in I}$, precision parameters $\varepsilon_j \doteq \|\nabla_x L(\tilde{x}^j, \lambda^j, \mu^j)\|_\infty$, where $\lambda^j = \tilde{\lambda}^j + \tilde{\rho}_j h(\tilde{x}^j)$ and $\mu^j = [\tilde{\mu}^j + \tilde{\rho}_j g(\tilde{x}^j)]_+$ for every $j \in I$, and any compact set $\mathcal{B}$ that contains every safeguarded multiplier. Then, since $\lim_{j\to\infty} \tilde{x}^j = x^*$, there is an instance of ALGENCAN that generates x^*. In order to see this, keep in mind that

$$\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k),$$

for every $k \in \mathbb{N}$.

The second case is when $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, which means x^* satisfies the KKT conditions since in this case $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded and $V^k \to 0$ ensures the fulfilment of the complementarity conditions. Denote its associated Lagrange multipliers by $\tilde{\lambda}$ and $\tilde{\mu}$. Then, choose any compact set $\mathcal{B}$ such that $(\tilde{\lambda}, \tilde{\mu}) \in \mathcal{B}$ and define $(\bar{\lambda}^k, \bar{\mu}^k) \doteq (\tilde{\lambda}, \tilde{\mu})$, $\rho_1 \doteq 1$, and $x^k \doteq x^*$, for every $k \in \mathbb{N}$. Thus, there is an instance of ALGENCAN that generates x^* in this case, as well.

The converse is trivial, since $V^k \to 0$ independently on the boundedness of $\{\rho_k\}_{k\in\mathbb{N}}$. □

In light of Theorem 1, AL-AKKT provides an ideal comparison tool between ALGENCAN and some of its variants, as well as a proper language for analysing the effects of some particular choices of parameters on the output of the algorithm. This is the main focus of Section 3. Before that, we discuss from the sequential optimality conditions perspective, some other interesting properties of ALGENCAN that are very often difficult to perceive with the algorithmic language.

2.2 AL-AKKT is necessary for optimality

A common approach for building the convergence theory of an algorithm is to ground it on some universal necessary optimality condition, that is, a condition that is independent of the algorithm (whether it is of the sequential type or not). However, such independence does not add any value for the algorithm at all and it may even turn into a limitation for its convergence theory. The following theorem shows that AL-AKKT can be interpreted as a necessary optimality condition, but since it is also an exact description of the augmented Lagrangian method, it provides the best possible convergence theory for it.

Theorem 2. *Every local minimizer x^* of* (NLP) *is an AL-AKKT point, regardless of the choice of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$.*

Proof. Let x^* be a local minimizer of (NLP). Then, there is a $\delta > 0$ such that x^* is the unique global solution of the localized problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) + 1/2\|x - x^*\|_2^2 \\ \text{subject to} & h(x) = 0, \quad g(x) \leqslant 0, \quad \|x - x^*\|_2 \leqslant \delta. \end{array} \tag{4}$$

The proof relies on arguments that are similar to the ones used in [22, Theorem 5.2]. In summary, the idea is to apply Algorithm 1 to (4), but assuming that we are able to compute a global solution of each subproblem in *Step 1* under the constraint $\|x - x^*\|_2 \leqslant \delta$, instead of only minimizing the correspondent augmented Lagrangian (1) up to first-order stationarity. Then, for each $k \in \mathbb{N}$, let x^k be a global solution of the augmented subproblem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) + \frac{1}{2}\|x - x^*\|_2^2 + \frac{\rho_k}{2}\sum_{i=1}^m \left[\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x)\right]^2 + \frac{\rho_k}{2}\sum_{j=1}^p \left[\frac{\bar{\mu}_j^k}{\rho_k} + g_j(x)\right]_+^2 \\ \text{subject to} & \|x - x^*\|_2 \leqslant \delta. \end{array} \tag{5}$$

Furthermore, the choice in *Step 3* can be made to match any sequence $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$ chosen a priori, highlighting that our conclusions are independent of this choice.

Let us show that $\{x^k\}_{k\in\mathbb{N}}$ converges to x^*. By the optimality of x^k and the feasibility of x^* we have, for all k, that

$$\begin{aligned} f(x^k) + \frac{1}{2}\|x^k - x^*\|_2^2 + \frac{\rho_k}{2}\sum_{i=1}^m \left[\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k)\right]^2 + \frac{\rho_k}{2}\sum_{j=1}^p \left[\frac{\bar{\mu}_j^k}{\rho_k} + g_j(x^k)\right]_+^2 \\ \leqslant f(x^*) + \sum_{i=1}^m \frac{(\bar{\lambda}_i^k)^2}{2\rho_k} + \sum_{j=1}^p \frac{(\bar{\mu}_j^k)^2}{2\rho_k}. \end{aligned} \tag{6}$$

Then, for all k, we have

$$\left(\frac{\bar{\lambda}^k}{\rho_k} + h(x^k)\,,\, \left[\frac{\bar{\mu}^k}{\rho_k} + g(x^k)\right]_+\right) = \left(\frac{\bar{\lambda}^k}{\rho_k}\,,\, \frac{\bar{\mu}^k}{\rho_k}\right) + \tilde{V}^k,$$

where $\tilde{V}^k \doteq (h(x^k), \max\{g(x^k), -\bar{\mu}^k/\rho_k\})$. Observe that $|\tilde{V}^k_\ell| = |V^k_\ell|$ for all $k \in \mathbb{N}$ and $\ell = 1, \ldots, m+p$. Thus, by (6), we get that

$$f(x^k) + \frac{1}{2}\|x^k - x^*\|_2^2 + \sum_{i=1}^{m} \bar{\lambda}^k_i \tilde{V}^k_i + \sum_{j=1}^{p} \bar{\mu}^k_j \tilde{V}^k_j + \frac{\rho_k}{2}\|\tilde{V}^k\|_2^2 \leqslant f(x^*). \tag{7}$$

Now, let us take an accumulation point $\bar{x}$ of $\{x^k\}_{k\in\mathbb{N}}$. There are two cases to consider:

- if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, then, by *Step 2* we have $\|\tilde{V}^k\|_2 = \|V^k\|_2 \to 0$ and therefore $\bar{x}$ is feasible. From (7), we have $f(\bar{x}) + (1/2)\|\bar{x} - x^*\|_2^2 \leqslant f(x^*)$. The optimality of x^* and the feasibility of $\bar{x}$ imply $\bar{x} = x^*$;
- if $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ then dividing (7) by ρ_k and taking limits lead us to obtain $\|V^k\|_2 = \|\tilde{V}^k\|_2 \to 0$, and thus $\bar{x}$ is feasible. Again by (7), we have $f(\bar{x}) + (1/2)\|\bar{x} - x^*\|_2^2 \leqslant f(x^*)$ and hence $\bar{x} = x^*$.

Thus, $x^k \to \bar{x} = x^*$ and $\|x^k - x^*\|_2 < \delta$ for all k large enough. Then, the optimality conditions for the penalized problem (5) at x^k give us

$$\nabla f(x^k) + \nabla h(x^k)[\bar{\lambda}^k + \rho_k h(x^k)] + \nabla g(x^k)[\bar{\mu}^k + \rho_k g(x^k)]_+ = x^* - x^k \to 0.$$

We also proved that $V^k \to 0$ regardless of whether $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded or not, concluding the proof. □

By Theorem 1, it is possible to say that the kind of convergence analysis provided by AL-AKKT via Theorem 2 is stronger than all known previous results regarding both: optimality conditions in the form "KKT or not-CQ" and sequential conditions. Indeed, every necessary condition that supports Algencan must be satisfied by all of its feasible limit points, and consequently implied by AL-AKKT. A detailed discussion on this topic can be found in Section 4, where we characterize the weakest CQ needed to establish convergence of Algencan to KKT points, by means of AL-AKKT.

Remark 2. *Similarly to the way Lagrange multipliers attest optimality of a KKT point in some sense, the sequences provided by an execution of* Algencan *can be seen as optimality certificates of an AL-AKKT point. In order to see this, keep in mind that even though such certificates were generated by a regularized variant of the method in the proof of Theorem 2, they are also valid outputs of* Algencan *with a suitable choice of parameters.*

We highlight that Theorems 1 and 2 also mean that the mere fact of being a feasible limit point of Algencan is itself a necessary optimality condition, which to the best of our knowledge, is novelty. While the usual convergence statement of Algencan tells us that the method is expected to converge to a local minimizer under some conditions, Theorem 1 complements it by stating that every local minimizer is likely to be found. This attribute should not be ignored in nonconvex problems with multiple local minima, since some local minimizers may be more interesting than the others (for instance, the global minimizer, when it does exist). In fact, whether this is an advantage for the method may be a situational issue, but the existence of some solutions that can be avoided by the method without specifying it should always raise a red flag. For instance, this is the case of an augmented Lagrangian for Generalized Nash Equilibrium problems. See the discussion in [24].

Remark 3. *Every KKT point is AL-AKKT as well. The reasoning is similar to the second case of the proof of Theorem 1: if x^* is KKT, let us say, with associated Lagrange multipliers $\tilde{\lambda}$ and $\tilde{\mu}$, then taking $x^k \doteq x^*$, $\rho_k \doteq k$, $\bar{\lambda}^k \doteq \tilde{\lambda}$, and $\bar{\mu}^k \doteq \tilde{\mu}$, for every $k \in \mathbb{N}$, is enough to conclude that x^* is also AL-AKKT.*

Remark 4. *If x^* is an AL-AKKT point and $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded (in particular, if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded), then it is a KKT point. In fact, in this case any accumulation point of $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ serves as Lagrange multipliers, and $V^k \to 0$ ensures the complementarity condition.*

A consequence of Theorem 2 and Remark 3 is that every possible point of interest for (NLP) is in the range of convergence of ALGENCAN.

2.3 About inexactly solving the subproblems

Note that *Step 1* only requires x^k to be an approximate stationary point of the Lagrangian function, $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$. This is done in order to maintain some degree of loyalty to what happens in practice, since either way x^k is likely to be obtained by some iterative method for solving an unconstrained subproblem, that declares convergence when the Lagrangian residue is small enough. Thus, we stress that we do not require any use of global methods for solving the subproblems, and similarly, no convexity assumptions are enforced.

However, it is natural to expect that the introduction of any kind of error tolerance might have negative influence over the outcome of the method. In [35] the authors study five *relaxation methods* for *mathematical problems with equilibrium constraints*, in particular, they compare the effects of forcing exactness on the subproblems (that is, an analogue of setting $\varepsilon_k = 0$ for every k in ALGENCAN) with the effects of allowing the subproblems to be solved inexactly. One of those methods was proven to keep the same convergence properties regardless of the exactness of the subproblems, whereas the others presented a strictly worse notion of convergence with inexactly solved subproblems.

Surprisingly, concerning ALGENCAN (via AL-AKKT), forcing exact stationarity by means of setting $\varepsilon_k = 0$ for every k may lead to unexpected and possibly undesired results, such as the exclusion of some local/global minimizers from the range of convergence of the method. The following example illustrates this fact:

Example 1. *In $\mathbb{R}^2$, consider the minimization problem:*

$$\text{minimize } -x_2 \quad \text{subject to} \quad x_2^2 \leqslant 0, \quad x_1^2 x_2^2 \leqslant 0.$$

Clearly, $x^* \doteq (1,0)$ is a global minimizer. Now, assume that there exists a sequence of stationary points of the Lagrangian $x^k \doteq (x_1^k, x_2^k)$ converging to x^* for some $\bar{\mu}_1^k$, $\bar{\mu}_2^k \in \mathbb{R}_+$. Thus, $\nabla L_{\rho^k,\bar{\mu}^k}(x^k)$ is equal to

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} + (\bar{\mu}_1^k + \rho_k (x_2^k)^2)_+ \begin{bmatrix} 0 \\ 2x_2^k \end{bmatrix} + (\bar{\mu}_2^k + \rho_k (x_1^k)^2 (x_2^k)^2)_+ \begin{bmatrix} 2x_1^k (x_2^k)^2 \\ 2x_2^k (x_1^k)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

From the first line of the expression above, we see that $2x_1^k(x_2^k)^2 = 0$ or $(\bar{\mu}_2^k + \rho_k(x_1^k)^2(x_2^k)^2) = 0$, for all $k \in \mathbb{N}$. In any case, since $x_1^k \to 1$ we get $x_2^k = 0$ for all k, contradicting the second expression.

Hence, the notion of convergence provided by Theorems 1 and 2 can be weakened under exactness, in contrast with our previous discussion. Moreover, Example 1 suggests that we should raise our caution for choosing the precision parameters ε_k, for the behavior of the method due to this choice is not trivially predictable. We refer to [18, 27] for some efficient strategies for choosing ε_k.

Similarly, it is also possible to prove that the augmented Lagrangian scheme can be incompatible with feasible sequences. In fact, forcing $V^k = 0$ for all k may not only remove minimizers from the range of convergence of ALGENCAN, but can also induce complete failure in the method. The following example illustrates this fact:

Example 2. *In $\mathbb{R}$, consider the minimization problem:*

$$\text{minimize } x_1 \quad \text{subject to} \quad x_1^2 \leqslant 0.$$

Clearly $x^* \doteq 0$ is the only global minimizer. Now, assume that there exists a sequence of exact admissible points x^k converging to x^* and multipliers $\bar{\mu}^k \in \mathbb{R}_+$, such that $1 + [\bar{\mu}^k + \rho_k(x_1^k)^2]_+ x_1^k \to 0$. Forcing $V^k = 0$ we get that $x_1^k = 0$, which is a contradiction with $1 + [\bar{\mu}^k + \rho_k(x_1^k)^2]_+ x_1^k \to 0$.

Note that this issue does not depend on the choice of $\bar{\mu}^k$. In fact, it is well-known that the feasibility of the limit point is a paramount issue for numerical methods based on penalty approaches, which is a huge contrast with *interior-point* and *active-set* methods, where feasibility is always maintained. Example 2 indicates that any attempt of resolving this issue by forcing feasibility in ALGENCAN may actually hinder its theoretical convergence.

In some situations, one may be interested in computing the Lagrange multipliers associated with a KKT point. Even though the KKT point itself can always be found by ALGENCAN, the choice of parameters may have direct influence on whether its Lagrange multipliers are computed along with it or not. The example below illustrates a scenario where a KKT point is found by ALGENCAN, but its Lagrange multiplier is not.

Example 3. *In $\mathbb{R}^2$, consider the minimization problem:*

$$\textit{minimize } -x_1^2 \quad \textit{subject to} \quad x_1^2 x_2 = 0.$$

Observe that $x^* \doteq (0,0)$ is a KKT point whose set of multipliers is the whole $\mathbb{R}$. Consider $x_1^k \doteq 1/k$, $x_2^k \doteq x_1^k$, $\rho_k \doteq (2x_1^k x_2^k)^{-2}$ and $\bar{\lambda}^k \doteq 0$. For this choice, we see that

$$\begin{bmatrix} -2x_1^k \\ 0 \end{bmatrix} + \rho_k (x_1^k)^2 x_2^k \begin{bmatrix} 2x_1^k x_2^k \\ (x_1^k)^2 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

However, observe that the approximate multiplier diverges:

$$\lambda^k \doteq \rho_k (x_1^k)^2 x_2^k = \frac{1}{2x_2^k} \to \infty.$$

Thus, since $\{(x_1^k, x_2^k)\}_{k\in\mathbb{N}}$ is a valid AL-AKKT sequence certified by $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, we can suppose without loss of generality, by the proof of Theorem 1, that the whole sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\lambda^k\}_{k\in\mathbb{N}}$ are valid outputs of ALGENCAN *where x^* is found, but its Lagrange multiplier is not.*

The next section deepens the discussion about implementation details of ALGENCAN by formally comparing some modifications and parameter choices that are often considered in practice, by means of AL-AKKT.

3 Comparing some augmented Lagrangian variants

There are various traits that distinguish ALGENCAN from other similar penalty-based algorithms, such as the *classical external penalty method* [42, Framework 17.1] and the *shifted external penalty method* [38, Algorithm 3.1]. However, a theoretical comparison among them in terms of their limit points has not been made yet, to the best of our knowledge. Nevertheless, ALGENCAN is often preferred over such variants since it allows for ρ_k to remain bounded under some circumstances, which indicates that it is numerically more stable than the others. In this section, we address this question via AL-AKKT. We also investigate the use of safeguarding and the effects of employing a different penalty parameter for each constraint.

3.1 The shifted external penalty method

The shifted external penalty method is a variant of ALGENCAN without the admissibility control for controlling the growth of ρ_k, that is, where *Step 2* is replaced by *Step 2-b*. In this section, we call this variant SHIFTED-EP.

In practice, ALGENCAN is expected to be numerically more stable than SHIFTED-EP due to more control on the growth of ρ_k. In theoretical terms, however, we see that there is no difference between them, which is somewhat surprising.

In order to make our argument clear, we present an alternative characterization of AL-AKKT:

Proposition 1. *A feasible point x^* is AL-AKKT if, and only if, there exist some sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that $\nabla_x L(x^k, \lambda^k, \mu^k) \to 0$, where $\lambda^k \doteq \bar{\lambda}^k + \rho_k h(x^k)$ and $\mu^k \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$.*

Algorithm 2 SHIFTED-EP

Input: A parameter $\gamma > 1$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial penalty parameter $\rho_1 > 0$, a compact set $\mathcal{B}$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Find an approximate stationary point x^k of $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\|\nabla L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x^k)\|_\infty \leqslant \varepsilon_k.$$

Step 2-b *(Updating the parameters):* Take ρ_{k+1} such that $\rho_{k+1} \geqslant \gamma\rho_k$;

Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k+1$, and go to Step 1.

Proof. If x^* is AL-AKKT, then there are sequences $\{x_0^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_{k,0}\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, $\{\bar{\lambda}_0^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, and $\{\bar{\mu}_0^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that (2) holds. If $\rho_{k,0} \to \infty$, there is nothing left to prove. Otherwise x^* satisfies the KKT conditions (Remark 4) with, let us say, multipliers $\tilde{\lambda}$ and $\tilde{\mu}$. Consequently, due to Remark 3, x^* is AL-AKKT associated with the new sequences defined by $x^k \doteq x^*$, $\rho_k \doteq k(\to \infty)$, $\bar{\lambda}^k \doteq \tilde{\lambda}$, and $\bar{\mu}^k \doteq \tilde{\mu}$, for every $k \in \mathbb{N}$. The converse is straightforward from the feasibility of x^* along with the fact $V^k \to 0$ when $\rho_k \to \infty$. □

This characterization turns out to be simpler than Definition 1 since it does not rely on the computation of V^k. However, Definition 1 is not completely replaceable because when (NLP) is convex, the augmented Lagrangian is guaranteed to globally converge with a fixed value of $\rho > 0$, hence the admissibility criterion of *Step 2* of ALGENCAN is always satisfied when k is large enough and $\{\rho_k\}_{k\in\mathbb{N}}$ is always bounded [46]. Even so, the characterization of Proposition 1 is strongly related to SHIFTED-EP, which allows us to conclude the following:

Theorem 3. *For each AL-AKKT point x^*, there is an instance of* SHIFTED-EP *such that its output sequence $\{x^k\}_{k\in\mathbb{N}}$ has x^* as an accumulation point. Conversely, every feasible accumulation point generated by* SHIFTED-EP *is AL-AKKT.*

Proof. Analogous to Theorem 1, in view of Proposition 1. □

However, the same conclusion is not necessarily valid for the classical external penalty method, which is the main focus of the next section.

3.2 The classical external penalty method

The classical external penalty method, which we call CLASSICAL-EP in this section, is precisely SHIFTED-EP with $\mathcal{B} = \{0\}$. In the previous section, we proved that ALGENCAN is equivalent to SHIFTED-EP in terms of their feasible limit points, so it is natural to question whether the same holds true for CLASSICAL-EP. The answer might seem obvious, for it is tempting to conclude that ALGENCAN and CLASSICAL-EP behave similarly, just because the shifts $\bar{\lambda}^k/\rho_k$ and $\bar{\mu}^k/\rho_k$ of ALGENCAN are small when ρ_k is large, but this is not entirely correct. The following example shows that the collection of all possible limit points of ALGENCAN contains, but may not be contained in the set of limit points of CLASSICAL-EP, and these extra points may be infeasible, even when $\rho_k \to \infty$. This is not necessarily a negative trait since we are mainly interested in feasible points. The purpose of the example is to emphasize that even when the shift is small, minimizing a shifted penalty function is not the same as minimizing the original penalty function.

Example 4. *Consider the following problem*

$$\textit{minimize } -x \quad \textit{subject to} \quad \sin(x) = 0, \quad \cos(x) = 0,$$

which is infeasible, which means that the admissibility test of Step 2 of ALGENCAN *will succeed only a finite number of times, so $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. In this scenario, a point x^* can be reached by*

ALGENCAN *when there is some sequence* $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ *and a bounded* $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ *such that*

$$-1 + \cos(x^k)(\bar{\lambda}_1^k + \rho_k \sin(x^k)) - \sin(x^k)(\bar{\lambda}_2^k + \rho_k \cos(x^k)) \to 0, \tag{8}$$

which holds if, and only if, $\cos(x^k)\bar{\lambda}_1^k - \sin(x^k)\bar{\lambda}_2^k \to 1$. *This is never satisfied when* $\bar{\lambda}^k = 0$, *so* CLASSICAL-EP *will never converge when applied to this problem. However, for every* $x^* \in \mathbb{R}$, *it is always possible to choose* $\bar{\lambda}^k$ *such that* x^* *is reached by* ALGENCAN.

Knowing that the behaviors of ALGENCAN and CLASSICAL-EP may differ when taking infeasible points into consideration, what is left is to compare their feasible limit points, but this is also not trivial, and a careful analysis must be done. A direct consequence of Theorem 2 is that the set of common limit points between ALGENCAN and CLASSICAL-EP contains at least every local minimizer. Next, we show that the same conclusion is also true for KKT points.

Theorem 4. *If* x^* *is a KKT point, there is some instance of* CLASSICAL-EP *such that its output sequence* $\{x^k\}_{k\in\mathbb{N}}$ *has* x^* *as an accumulation point.*

Proof. Let x^* be a KKT point of (NLP), and consider the sets

$$K(x) \doteq \{\nabla h(x)\lambda + \nabla g(x)\mu \mid \mu \geqslant 0,\ \mu_j = 0 \text{ if } g_j(x^*) < 0\} \tag{9}$$

and

$$K(x, \rho) \doteq \{\nabla h(x)[\rho h(x)] + \nabla g(x)[\rho g(x)]_+\}, \tag{10}$$

where $\rho > 0$. Clearly, the fact that x^* is a KKT point is equivalent to $-\nabla f(x^*) \in K(x^*)$. Note that x^* is a limit of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by CLASSICAL-EP if $x^k \to x^*$ and

$$\nabla f(x^k) + \nabla h(x^k)[\rho_k h(x^k)] + \nabla g(x^k)[\rho_k g(x^k)]_+ \to 0$$

for a certain sequence $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. In other words, the classical external penalty is capable of reaching x^* if

$$-\nabla f(x^*) \in \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho).$$

So, in order to prove that x^* can be reached by CLASSICAL-EP, it is sufficient to prove that

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

Let $\mathcal{T}(x^*)$ be the tangent cone to the feasible set of (NLP) at x^*, and let $\mathcal{L}(x^*)$ be its linearized cone. It is always true that $\mathcal{T}(x^*) \subset \mathcal{L}(x^*)$, which implies $K(x^*) = \mathcal{L}(x^*)^\circ \subset \mathcal{T}(x^*)^\circ$. Now, given $w^* \in K(x^*)$, Lemma 4.3 of [15] ensures that there are sequences $\{w^k\}_{k\in\mathbb{N}} \to w^*$ and $\{\tilde{x}^k\}_{k\in\mathbb{N}} \to x^*$ such that

$$w^k = \nabla h(\tilde{x}^k)[kh(\tilde{x}^k)] + \nabla g(\tilde{x}^k)[kg(\tilde{x}^k)]_+.$$

As $w^k \in K(\tilde{x}^k, k)$ for all k, we have $w^* \in \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)$, concluding the proof. □

To the best of our knowledge, the fact that CLASSICAL-EP may converge to any KKT point (Theorem 4) is new in the literature. As an immediate consequence of this fact, we see that if $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ remains bounded in Definition 1 (in particular, if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded), then ALGENCAN and CLASSICAL-EP are indistinguishable in terms of the quality of their limit points. Formally:

Corollary 1. *If* x^* *is an AL-AKKT point such that* $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ *is bounded, then there is an associated AL-AKKT sequence* $\{x^k\}_{k\in\mathbb{N}}$ *with* $(\bar{\lambda}^k, \bar{\mu}^k) \doteq (0, 0)$ *for all* k. *That is, in this case every feasible limit point of* ALGENCAN *can be reached by* CLASSICAL-EP *and vice-versa.*

Proof. Since $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded, x^* is a KKT point (Remark 4). So, the conclusion follows from Theorem 4. □

It is still not clear, for an arbitrary choice of $\bar{\lambda}^k$ and $\bar{\mu}^k$, whether every limit point of ALGENCAN is reachable by an instance of CLASSICAL-EP or not. Nevertheless, we conjecture they are equivalent regarding their feasible limit points. Some facts that support our belief are discussed in Section 3.4 and in Section 4, within Theorem 6 and Corollary 2.

3.3 On multiplier updates and safeguards

We presented ALGENCAN with a safeguarding procedure, where $\bar{\lambda}^k$ and $\bar{\mu}^k$ are taken from a compact set, but the original augmented Lagrangian algorithm proposed by Hestenes and Powell [30, 43] employs a different strategy. In their work, they use

C1. $\bar{\lambda}^{k+1} \doteq \bar{\lambda}^k + \rho_k h(x^k)$ and $\bar{\mu}^{k+1} \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$ for every $k \in \mathbb{N}$,

that is, $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \doteq (\lambda^k, \mu^k)$ for every k, regardless of $\mathcal{B}$, which means $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ are not necessarily bounded.

Clearly, there is no difference between employing **C1** and safeguarding when $\mathcal{B}$ is large enough and the sequences $\{\lambda^k\}_{k\in\mathbb{N}}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ converge to some Lagrange multipliers $\tilde{\lambda}$ and $\tilde{\mu}$, respectively. This holds, for instance, when $(\bar{x}, \tilde{\lambda}, \tilde{\mu})$ is a KKT point that satisfies the *second-order sufficient optimality condition* (SOSC) and the augmented Lagrangian scheme starts sufficiently close to $(x^*, \tilde{\lambda}, \tilde{\mu})$, with a penalty parameter ρ_1 large enough and ε_k appropriately controlled (see [27, Theorem 3.4]). On the other hand, without such assumptions, the behaviour of ALGENCAN with **C1** can be very different from ALGENCAN with safeguards, as it was shown in [36] by means of a simple example, with a unique minimizer that was also a KKT point. In their example, the safeguarded method was proven to generate sequences whose feasible limit points are exactly the minimizer of the problem, whilst the method that employs **C1** was proven to be unable to converge to it. Hence, in terms of reliability, using safeguarded multipliers is better than using **C1**, at least theoretically.

Continuing the discussion from [36], we are led to investigate the effects of removing safeguards from ALGENCAN, but without limiting the choices of $\bar{\lambda}^k$ and $\bar{\mu}^k$ to **C1**, that is, we consider $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$. First, note that a modest control over the parameters, such as

$$\frac{\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2}{\rho_k} \to 0 \tag{11}$$

is enough to establish a reasonable convergence theory for ALGENCAN with $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$, which goes in the exact same lines as the proofs of Theorems 1 and 2. In order to highlight this fact, we state it in a theorem environment as follows:

Theorem 5. *Let x^* be a local minimizer of* (NLP)*. Then, x^* is reachable by* ALGENCAN *with $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$, regardless of the choices of the sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, and $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, as long as* (11) *holds.*

Note that (11) holds whenever $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ are bounded and $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, or when $\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2 = O(\rho_k^\beta)$, for every $\beta \in (0,1)$. Thus, it allows a certain freedom in the growth of the approximate multipliers. However, the following example shows that it is not possible to relax (11) even further, for instance assuming $\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2 = O(\rho_k)$, without losing the property described in Theorem 5.

Example 5. *In $\mathbb{R}^2$, consider the minimization problem:*

$$\textit{minimize } -x_2 \quad \textit{subject to} \quad x_1^2 x_2 \leqslant 0.$$

Note that $x^* \doteq (1,0)$ is a local minimizer. Assume that there exists a sequence $x^k \doteq (x_1^k, x_2^k)$ with $x^k \to x^*$ and some approximate multiplier $\bar{\mu}^k$ with $\bar{\mu}^k/\rho_k \geqslant \alpha$, where $\alpha > 0$, for all k large enough, with $\rho_k \to \infty$, such that

$$\nabla L_{\rho,\bar{\mu}}(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} + [\bar{\mu}^k + \rho_k (x_1^k)^2 (x_2^k)]_+ \begin{bmatrix} 2x_1^k x_2^k \\ (x_1^k)^2 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus, $[\bar{\mu}^k + \rho_k (x_1^k)^2 (x_2^k)]_+ \to 1$. Then, $[(\bar{\mu}^k/\rho_k) + (x_1^k)^2 (x_2^k)]_+ \to 0$ and this implies $\bar{\mu}^k/\rho_k \to 0$, which is a contradiction.

3.4 Independent penalty parameters

Here we consider the possibility of different penalty parameters ρ's, one for each constraint. At first glance, this modification appears to make Algencan more flexible, but it is of common sense that the use of different ρ's leads to sightly worse computational results (see [2]). In this section we show in a formal way that using a common ρ is indeed better than using different ρ's. We refer to the variant of Algencan with independent penalty parameters by the name Sep-Algencan.

Algorithm 3 Sep-Algencan

Input: Some arrays of parameters $\tau \in (0,1)^{m+p}$ and $\gamma \in (1,\infty)^{m+p}$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial array of penalty parameters $\rho^1 \in \mathbb{R}^{m+p}_{++}$, a compact set $\mathcal{B} \subseteq \mathbb{R}^m \times \mathbb{R}^p_+$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho^k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\|\nabla L_{\rho^k,\bar{\lambda}^k,\bar{\mu}^k}(x^k)\|_\infty \leqslant \varepsilon_k.$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \left(h(x^k), \min\left\{ -g_1(x^k), \frac{\bar{\mu}_1^k}{\rho_{m+1}^k} \right\}, \ldots, \min\left\{ -g_p(x^k), \frac{\bar{\mu}_p^k}{\rho_{m+p}^k} \right\} \right). \tag{12}$$

Then, for each $\ell \in \{1, \ldots, m+p\}$,

a. If $k = 1$ or $|V_\ell^k| \leqslant \tau_\ell |V_\ell^{k-1}|$, set $\rho_\ell^{k+1} \doteq \rho_\ell^k$;

b. Otherwise, take some ρ_ℓ^{k+1} such that $\rho_\ell^{k+1} \geqslant \gamma_\ell \rho_\ell^k$.

Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k+1$ and go to Step 1.

In order to make a proper analysis, we deal with the following version of AL-AKKT with separate ρ's:

Definition 2. *We say that x^* is a* Sep-AL-AKKT *point if there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^{m+p}_{++}$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that*

$$\nabla_x L(x^k, \lambda^k, \mu^k) \to 0, \quad \text{and} \quad V^k \to 0$$

where $\lambda_i^k \doteq \bar{\lambda}_i^k + \rho_i^k h_i(x^k)$ for all $i = 1, \ldots, m$, $\mu_j^k \doteq [\bar{\mu}_j^k + \rho_{m+j}^k g_j(x^k)]_+$ for all $j = 1, \ldots, p$, and V^k is defined as in (12).

A procedure analogous to the proof of Theorem 1 is enough to conclude that every feasible limit point of an instance of Sep-Algencan must be Sep-AL-AKKT, and that every Sep-AL-AKKT point is an accumulation point of an instance of Sep-Algencan. Thus, the task of comparing Algencan with Sep-Algencan reduces to comparing Definitions 1 and 2.

Evidently, every AL-AKKT sequence is a Sep-AL-AKKT one, since the former is a particular case of the latter. However, the next example shows that the converse is not necessarily true.

Example 6. *Let us consider the problem*

$$\text{minimize } x_2 \quad \text{subject to} \quad x_1^2 x_2 = 0, \quad x_1 = 0,$$

adapted from [4, Example 3]. Using ρ_1 and ρ_2 for first and second constraints, respectively, we have

$$\nabla L_{\rho,\bar{\lambda}}(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + [\bar{\lambda}_1 + \rho_1 x_1^2 x_2] \begin{bmatrix} 2x_1 x_2 \\ x_1^2 \end{bmatrix} + [\bar{\lambda}_2 + \rho_2 x_1] \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{13}$$

It is straightforward to verify that the feasible point $x^* \doteq (0,0)$ *is Sep-AL-AKKT with*

$$\tilde{x}^k \doteq (1/k^2, -1/k^3), \quad \bar{\lambda}^k \doteq (0,0), \quad \tilde{\rho}^k \doteq (k^{11}, k).$$

Now we will show that x^* *is not AL-AKKT. As the gradient of the first constraint vanishes at* x^**, the bounded multiplier sequence* $\{\bar{\lambda}_1^k\}_{k\in\mathbb{N}}$ *does not matter in our analysis. Thus we omit it. Suppose that* $\{x^k\}_{k\in\mathbb{N}} \to x^*$ *is an AL-AKKT sequence with the associated sequence* $\{\rho^k\}_{k\in\mathbb{N}}$ *and* $\{\bar{\lambda}_2^k\}_{k\in\mathbb{N}}$*. From* (13) *with* $\rho_1^k = \rho_2^k \doteq \rho^k$ *we must have*

$$2\rho^k (x_1^k)^3 (x_2^k)^2 + \rho^k x_1^k + \bar{\lambda}_2^k \to 0 \quad \text{and} \quad 1 + \rho^k (x_1^k)^4 x_2^k \to 0. \tag{14}$$

From the second expression of (14)*, we have* $|\rho^k x_1^k| \to \infty$*, and then using the first expression,* $\rho^k x_1^k \,[2(x_1^k)^2(x_2^k)^2 + 1] + \bar{\lambda}_2^k \to 0$*, which leads to a contradiction.*

That is, methods that employ a common ρ for all constraints are better, at least regarding the quality of the limit points. Also, note that the counterpart of Classical-EP with different penalty parameters is equivalent to Sep-Algencan, which complements the discussion of Section 3.2. Indeed, if x^* satisfies Sep-AL-AKKT with sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$, $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, and $\{\rho^k\}_{k\in\mathbb{N}}$, we can define new parameters

$$\bar{\rho}_i^k \doteq \begin{cases} \frac{\bar{\lambda}_i^k}{h_i(x^k)} + \rho_i^k & \text{if } h_i(x^k) \neq 0 \\ \rho_i^k & \text{otherwise} \end{cases}, \quad \bar{\rho}_{m+j}^k \doteq \begin{cases} \frac{\bar{\mu}_j^k}{[g_j(x^k)]_+} + \rho_{m+j}^k & \text{if } [g_j(x^k)]_+ \neq 0 \\ \rho_{m+j}^k & \text{otherwise} \end{cases},$$

for all $i \in \{1,\dots,m\}$ and all $j \in \{1,\dots,p\}$, so that x^* is also a limit point of Classical-EP with different ρ's. An immediate consequence of this fact is that the property of being Sep-AL-AKKT is invariant to the choice of $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$.

4 Tightest global convergence theory of the safeguarded augmented Lagrangian method to KKT points

One of the signature features of sequential optimality conditions is the presence of a companion *strict constraint qualification* (SCQ), which is basically the weakest CQ that guarantees equivalence between a given sequential optimality condition and the KKT conditions (see [16] for details). They are useful when compared with classical CQs, such as MFCQ [37] and Abadie's CQ [1], for acting like a measurement tool for the strength of the sequential optimality condition associated with them.

Since AL-AKKT can be viewed as a sequential optimality condition, it also has a companion SCQ, which is the main focus of this section. Let us recall the cones (9) and (10), related to the KKT conditions and Classical-EP, respectively. For a feasible x^* and $\rho > 0$, we have

$$K(x) = \{\nabla h(x)\lambda + \nabla g(x)\mu \mid \mu \geqslant 0,\ \mu_j = 0 \text{ if } g_j(x^*) < 0\}$$

and

$$K(x,\rho) = \{\nabla h(x)[\rho h(x)] + \nabla g(x)[\rho g(x)]_+\}.$$

From the proof of Theorem 4, we have $K(x^*) \subset \limsup_{x\to x^*,\, \rho\to\infty} K(x,\rho)$. The reverse inclusion is not always true (since we will show in Section 4.1 that this property strictly implies Abadie's CQ). This gap is exactly what characterizes the SCQ associated with AL-AKKT.

Definition 3. *We say that a feasible point* x^* *satisfies the* AL-regularity *condition if*

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

We use the name "AL-regularity" in the same fashion as [16]. The following result is a formal statement that AL-regularity is indeed the SCQ associated with AL-AKKT.

Theorem 6. *Every AL-AKKT point satisfying the AL-regularity condition is KKT. Conversely, if an AL-AKKT point* x^* *is also KKT, for every objective function* f*, then* x^* *satisfies the AL-regularity condition.*

Proof. Let x^* be an AL-AKKT point with associated sequences $\{x^k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$. If $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, x^* is KKT independently of the presence of AL-regularity (see Remark 4). Thus, from now on suppose that $\rho_k \to \infty$. For simplicity, we denote

$$r(z, a, b) \doteq \nabla h(z)a + \nabla g(z)b.$$

Observe that

$$r(x^k, \bar{\lambda}^k + \rho_k h(x^k), [\bar{\mu}^k + \rho_k g(x^k)]_+) = r(x^k, \rho_k h(x^k), [\rho_k g(x^k)]_+) + r(x^k, \bar{\lambda}^k, \tilde{\mu}^k) \tag{15}$$

where $\tilde{\mu}^k := [\bar{\mu}^k + \rho_k g(x^k)]_+ - \rho_k g(x^k)_+ \geqslant 0$. Clearly, $r(x^k, \rho_k h(x^k), [\rho_k g(x^k)]_+) \in K(x^k, \rho_k)$ and, from the boundedness of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$, we also have, taking a subsequence if necessary, that $r(x^k, \bar{\lambda}^k, \tilde{\mu}^k)$ converges to some element in $K(x^*)$. Furthermore, observe that the left-hand side of (15) converges to $-\nabla f(x^k)$. Thus, by the validity of AL-regularity at x^*, we conclude that

$$-\nabla f(x^*) \in \left[\limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)\right] + K(x^*) \subset K(x^*) + K(x^*) = K(x^*),$$

that is, x^* is a KKT point.

Conversely, let $w^* \in \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)$. Then, there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{w^k\}_{k\in\mathbb{N}} \to w^*$ and $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ such that

$$w^k = r(x^k, \rho_k h(x^k), \rho_k g(x^k)_+) \in K(x^k, \rho_k)$$

for all k. Defining $f(x) = -(w^*)^T x$, the above expression gives

$$\nabla f(x^k) + \nabla h(x^k)[\rho_k h(x^k)] + \nabla g(x^k)[\rho_k g(x^k)_+] = -w^* + w^k \to 0.$$

Thus, we conclude that x^* is an AL-AKKT point with $(\bar{\lambda}^k, \bar{\mu}^k) = 0$ for all k. By hypothesis, it is also KKT. This implies $w^* = -\nabla f(x^*) \in K(x^*)$, and thus AL-regularity holds at x^*. □

An immediate use of AL-regularity is to give a partial answer to the question about the equivalence between Classical-EP and Algencan. Recall that the sequences generated by Classical-EP are exactly AL-AKKT sequences with $(\bar{\lambda}^k, \bar{\mu}^k) = 0$, for all k, and $\rho_k \to \infty$. Thus, in view of Theorem 6, we conclude that AL-regularity is also the weakest CQ that guarantees convergence of Classical-EP to KKT points.

Corollary 2. *Every AL-AKKT point with $(\bar{\lambda}^k, \bar{\mu}^k) \doteq 0$ for all k and $\rho_k \to \infty$ satisfying the AL-regularity condition is KKT. Conversely, if an AL-AKKT point x^* with $(\bar{\lambda}^k, \bar{\mu}^k) \doteq 0$ for all k and $\rho_k \to \infty$ is also KKT, for every objective function f, then x^* satisfies the AL-regularity condition.*

Proof. The first statement follows from Theorem 6. Conversely, the reader may notice that, in the proof of Theorem 6, we take $(\bar{\lambda}^k, \bar{\mu}^k) = 0$ for all k. □

The above discussion strongly suggests that the feasible accumulation points of sequences generated by Algencan and those generated by Classical-EP are exactly the same, however this should be investigated thoroughly in a future work.

We strongly emphasize that AL-regularity has a deeper meaning concerning the augmented Lagrangian because it describes the weakest CQ qualification required to prove convergence of Algencan to KKT points. What follows is a formal statement of this fact.

Corollary 3. *Let x^* be a feasible accumulation point of a sequence generated by* Algencan*. If x^* satisfies AL-regularity then it also satisfies the KKT conditions.*

Proof. This is a direct consequence of Theorems 1 and 6. □

Thus, the best possible global convergence theory for Algencan is the one built around Definition 1, Theorems 1 and 2, and Corollary 3. That means the only way of obtaining a stronger theory is imposing additional conditions over *Step 1* of the method. For example, requiring x^k to satisfy some approximate second-order necessary condition, in the sense of [8], is likely to lead to better results.

4.1 Relations of AL-regularity with other constraint qualifications

From the perspective of sequential optimality conditions, a natural question that arises is about the relation between other SCQs and AL-regularity outside the context of Algencan. To the best of our knowledge, among the strongest sequential optimality conditions existent in the literature there are the *Positive AKKT* (PAKKT) condition and the *Approximate Gradient Projection* (AGP) condition, presented in [4] and [40], respectively. This means their associated SCQs are among the weakest possible. We are interested in both of them because they are independent of each other [4].

The AGP condition from [40] holds at a feasible point x^* when there exists some sequence $\{x^k\}_{k\in\mathbb{N}} \to x^*$ such that

$$\|\text{proj}_{\Omega(x^k)}(x^k - \nabla f(x^k)) - x^k\|_\infty \to 0, \tag{16}$$

where

$$\Omega(x) \doteq \{z \in R^n \mid \nabla h(x)^T(z-x) = 0,\ \min\{0, g(x)\} + \nabla g(x)^T(z-x) \leqslant 0\},$$

which can be viewed as a linearization of the feasible set of (NLP) at $x \in \mathbb{R}^n$. Note that AGP does not explicitly require a Lagrange multiplier approximation to be verified. For this reason, it is suitable for proving convergence of numerical methods that do not provide multiplier approximations, such as *inexact restoration* methods [39]. Recently, however, it was proved that there is an equivalent version of AGP with Lagrange multipliers [3, Theorem 2.7].

According to [4, Definition 2.1], the PAKKT condition holds at a feasible point x^* of (NLP) when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\lambda^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^p$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^m_+$ such that $\nabla_x L(x^k, \lambda^k, \mu^k) \to 0$, $\min\{-g(x^k), \mu^k\} \to 0$, and additionally,

$$\lim_{k\to\infty} \frac{|\lambda_i^k|}{\delta^k} > 0 \Rightarrow \lambda_i^k h_i(x^k) > 0, \quad \text{and} \quad \lim_{k\to\infty} \frac{\mu_i^k}{\delta^k} > 0 \Rightarrow \mu_i^k g_i(x^k) > 0,$$

for all k sufficiently large, where $\delta^k \doteq \|(1, \lambda^k, \mu^k)\|_\infty$, for every $k \in \mathbb{N}$. It was born as an improvement for the AKKT condition where the sign of the multipliers are controlled, even the ones associated with equality constraints. The difference with AKKT becomes very evident when considering *complementarity constraints* in (NLP). With regard to Algencan, [4, Theorem 4.1] states that all of its feasible limit points must satisfy PAKKT with no additional assumption.

Now, since every AL-AKKT point is a feasible limit point of Algencan (Theorem 1), which in turn satisfies both AGP and PAKKT, it follows that:

Theorem 7. *AL-AKKT implies both AGP and PAKKT.*

Following [4, Definition 2.3], the SCQ associated with PAKKT, namely *PAKKT-regularity*, consists in an upper semicontinuity-like of the mapping

$$K^P(x, \alpha, \beta) \doteq \left\{ r(x, \lambda, \mu) \in K(x) \middle| \begin{array}{l} \lambda_i h_i(x) \geqslant \alpha \text{ if } |\lambda_i| > \beta\|(1, \lambda, \mu)\|_2 \\ \mu_j g_j(x) \geqslant \alpha \text{ if } \mu_j > \beta\|(1, \lambda, \mu)\|_2 \end{array} \right\},$$

that is, PAKKT-regularity holds at x^* when

$$\limsup_{x\to x^*,\, \alpha\downarrow 0,\, \beta\downarrow 0} K^P(x, \alpha, \beta) \subset K(x^*),$$

which is similar to Definition 3 in some sense. Furthermore, the SCQ associated with the AGP condition from [16, Definition 1], namely *AGP-regularity*, is characterized by the upper semicontinuity of the normal cone of $\Omega(x)$. That is, we say that a feasible point x^* is AGP-regular when

$$\limsup_{x\to x^*,\, \varepsilon\to 0} N_{\Omega(x)}(x+\varepsilon) \subset N_{\Omega(x^*)}(x^*).$$

In order to compare AL-regularity with the strict constraint qualifications associated with PAKKT and AGP, it suffices to keep in mind the following characterization, presented in [4, Theorem 2.4] and [16, Theorem 1]:

Theorem 8. *A feasible point x^* of* (NLP) *satisfies PAKKT-regularity (respectively, AGP-regularity) if, and only if, for every continuously differentiable objective function, the fact of x^* being PAKKT (respectively, AGP) implies that x^* is KKT.*

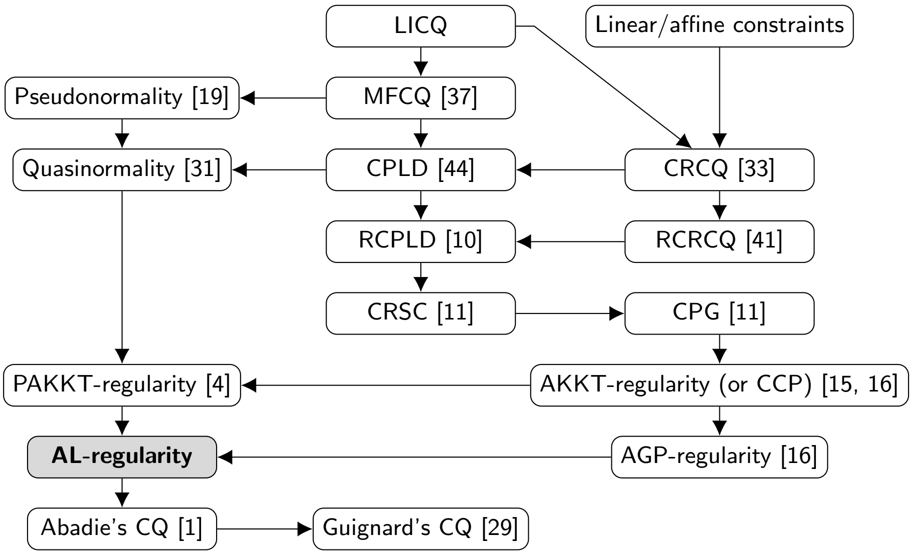

Figure 1: Updated landscape of constraint qualifications for standard nonlinear programming. AL-regularity and every more stringent CQ are associated with the global convergence of Algencan.

Hence, their relation with AL-regularity follows from Theorem 7:

Corollary 4. *Both AGP-regularity and PAKKT-regularity imply AL-regularity.*

Proof. Let x^* be a feasible point of (NLP) satisfying AGP-regularity. Then, for every continuously differentiable objective function we have that: if x^* is AGP, then it is also KKT. Hence, we also have that for every continuously differentiable objective function, if x is AL-AKKT, it is also AGP (due to Theorem 7) and, consequently, KKT. Now, it follows from Theorem 6 that x^* satisfies AL-regularity. Thus, AGP-regularity implies AL-regularity. Analogously, it is possible to prove that PAKKT-regularity also implies AL-regularity. □

It is important to note that the implications given by Corollary 4 are strict, since PAKKT and AGP are independent [4].

Finally, we present the relation between Abadie's CQ and AL-regularity. Let us denote the tangent cone to the feasible set of (NLP) at x^* by $\mathcal{T}(x^*)$, and its linearized cone by $\mathcal{L}(x^*)$. We say that Abadie's CQ holds at x^* if $\mathcal{T}(x^*) = \mathcal{L}(x^*)$. See [1].

Theorem 9. *AL-regularity implies Abadie's CQ.*

Proof. The proof is analogous to that of [15, Theorem 4.4], since for the sequences $\{x^{k,l(k)}\}_{k\in\mathbb{N}}$ and $\{w^{k,l(k)}\}_{k\in\mathbb{N}}$ considered in that proof, $w^{k,l(k)} \in K(x^{k,l(k)}, l(k))$ holds true for all k, and $\lim_{k\to\infty} l(k) = \infty$. □

The implication in Theorem 9 is strict, as the next example shows.

Example 7 (Abadie's CQ does not imply AL-regularity)**.** *As in [19, Example 7.3], let us consider the constraints in* $\mathbb{R}^2$

$$x_1^6 + x_2^3 \leqslant 0, \quad x_2 \leqslant 0$$

and the feasible point $x^* \doteq (0,0)$. *It is straightforward to verify that* $\mathcal{T}(x^*) = \mathcal{L}(x^*) = \mathbb{R} \times \mathbb{R}_-$, *and thus* x^* *conforms to Abadie's CQ. On the other hand, AL-regularity does not hold at* x^* *since* $K(x^*) = \{0\} \times \mathbb{R}_+$ *and, defining the sequences* $\rho_k \doteq k^{11}$ *and* $x^k \doteq (1/k, 0)$ *for all* $k \in \mathbb{N}$, *we see that* $K(x^*) \not\ni (6,0) \in \limsup_{x\to x^*,\, \rho\to\infty} K(x,\rho)$.

We summarize the relation among several known CQs from the literature and AL-regularity in Figure 1.

5 Conclusions

Over the last years, the convergence of ALGENCAN has been enhanced by means of the so-called sequential optimality conditions (see [3, 4, 16, 17]), and the best results so far are related to the independent sequential conditions PAKKT [4] and AGP [6]. As a rule, sequential optimality conditions must be indeed necessary for optimality, but not only that: they must imply the KKT conditions under some mild constraint qualifications, and there must be at least one relevant algorithm whose feasible limit points satisfy them, in order to illustrate its usability. In this paper, we defined a new sequential optimality condition, called AL-AKKT, which not only satisfies the three properties we just mentioned (Theorems 2, 6, and 1, respectively), but also completely characterizes all feasible limit points of ALGENCAN. Consequently, its associated strict constraint qualification characterizes the weakest possible constraint qualification under which ALGENCAN is guaranteed to converge to KKT points. In particular, since AL-AKKT strictly implies both conditions PAKKT and AGP (Theorem 7), Theorem 6 improves the global convergence result from [4, Theorem 4.1], as well as guarantees that ALGENCAN converges to AGP points.

From a practical point of view, we recall that there are many distinct variants of the augmented Lagrangian method, with potentially different performances depending on the problem they are applied to. In a real world application, one may find suitable to use different implementations of the method to solve the same problem and then select the best solution among their outputs, afterwards. But should this be impossible or inconvenient, the results of this paper can be taken into consideration for deciding, beforehand, which implementation might be the best for a general problem, regarding the quality of the solutions that it may return. Based on our findings, we believe that the most reasonable implementation of the augmented Lagrangian, that balances theory and practice, is characterized by:

- The use of projected (bounded, safeguarded) Lagrange multipliers in the underlying problems of the method, since this leads to solutions with the same quality as the pure external penalty method (see Corollary 1 and the discussion in Section 3.3);
- The use of a penalty parameter growth control, since it has a positive effect over the numerical stability of the method (see, for instance, [2]) without any drawback on its convergence theory (Theorem 3);
- The use of a single penalty parameter ρ for all constraints, as suggested in Section 3.4.

Besides, it is reasonable that the set $\mathcal{B}$ of projected multipliers be large, for this increases the likelihood of $\bar{\lambda}$ and $\bar{\mu}$ to converge to actual Lagrange multipliers (when they exist), hence avoiding unnecessary increments of ρ. See also [2, Section 5] and the book [22]. As a matter of fact, the most similar variant to what we just described is the ALGENCAN implementation.

As for the optimality condition AL-AKKT, we remark that it contrasts with other sequential conditions, which are meant for unifying convergence theories of different algorithms, since it is intrinsic to AL strategies. Nevertheless we believe our approach, that consists of characterizing a specialized sequential condition from an algorithm, may be useful for building convergence theories of other algorithms, for comparing some of their variants, and ultimately, for comparing different algorithms, from a theoretical point of view. For instance, in [14], the authors prove that the Newton–Lagrange method may not generate even AKKT sequences, which is the weakest form of sequential optimality condition in the literature. Moreover, some promising results were presented recently in [9] assessing the use of a stopping criterion for augmented Lagrangian methods based on a scaled KKT residual. This indicates that a convergence analysis of these algorithms based on an analogue of AL-AKKT should bring to light some interesting aspects of them. We also expect other algorithms to be analysed with the ideas introduced in this paper, together with extensions to other types of optimization frameworks [5, 6, 12, 13, 23, 24, 26, 28, 34, 45].

Acknowledgements

We would like to thank the reviewers of this paper for their many insightful suggestions.

References

[1] J. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear Programming*, pages 21–36. John Wiley, 1967.

[2] 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(4):1286–1309, 2007.

[3] R. Andreani, A. R. V. Cárdenas, A. Ramos, A. Ribeiro, and L. D. Secchin. On the convergence of augmented Lagrangian strategies for nonlinear programming. 2020. To appear in IMA Journal of Numerical Analysis.

[4] R. Andreani, N. Fazzio, M. Schuverdt, and L. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, 29(1):743–766, 2019.

[5] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin. Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report, 2019.

[6] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos. On optimality conditions for nonlinear conic programming. Technical report, 2020.

[7] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[8] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017.

[9] R. Andreani, G. Haeser, M. L. Schuverdt, L. D. Secchin, and P. J. S. Silva. On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. *To appear in Mathematical Programming Computation*, 2021.

[10] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1):255–273, 2012.

[11] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012.

[12] 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.

[13] R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 2018.

[14] R. Andreani, J. M. Martínez, and L. T. Santos. Newton's method may fail to recognize proximity to optimal points in constrained optimization. *Mathematical Programming*, 160(1):547–555, 2016.

[15] 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(1):96–110, 2016.

[16] 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(3):693–717, 2018.

[17] 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.

[18] J. S. Arora, A. I. Chahande, and J. K. Paeng. Multiplier methods for engineering optimization. *International Journal for Numerical Methods in Engineering*, 32(7):1485–1525, 1991.

[19] 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.

[20] E. G. Birgin, R. A. Castillo, and J. M. MartÍnez. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. *Computational Optimization and Applications*, 31(1):31–55, 2005.

[21] E. G. Birgin and J. M. Martínez. Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization. *Computational Optimization and Applications*, 51(3):941–965, 2012.

[22] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[23] L. Bueno, G. Haeser, F. Lara, and F. Rojas. An augmented Lagrangian method for quasi-equilibrium problems. *Computational Optimization and Applications*, 76:737–766, 2020.

[24] L. Bueno, G. Haeser, and F. Rojas. Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications. *SIAM Journal on Optimization*, 29(1):31–54, 2019.

[25] F. E. Curtis, H. Jiang, and D. P. Robinson. An adaptive augmented Lagrangian method for large-scale constrained optimization. *Mathematical Programming*, 152:201–245, 2015.

[26] M. Feng and S. Li. An approximate strong KKT condition for multiobjective optimization. *TOP*, 26(3):489–509, 2018.

[27] D. Fernandez 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.

[28] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM Journal on Optimization*, 30(2):1067–1093, 2020.

[29] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[30] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

[31] M. R. Hestenes. *Optimization Theory: The Finite Dimensional Case*. John Wiley & Sons, New York, 1975.

[32] 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.

[33] R. Janin. Directional derivative of the marginal function in nonlinear programming. In A. V. Fiacco, editor, *Sensitivity, Stability and Parametric Analysis*, volume 21 of *Mathematical Programming Studies*, pages 110–126. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[34] C. Kanzow, A. B. Raharja, and A. Schwartz. Sequential optimality conditions for cardinality-constrained optimization problems with applications. Technical report, 2021.

[35] C. Kanzow and A. Schwartz. The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. *Mathematics of Operations Research*, 40(2):253–275, 2015.

[36] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters*, 45(6):598–603, 2017.

[37] 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(1):37–47, 1967.

[38] J. M. Martínez. Otimização prática usando o lagrangiano aumentado (in portuguese). Technical report, 2006.

[39] J. M. Martínez and E. A. Pilotta. Inexact-restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, 104(1):135–163, 2000.

[40] J. M. Martínez and B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003.

[41] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[42] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.

[43] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, New York, 1969.

[44] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[45] A. Ramos. Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization Methods and Software*, pages 1–37, 2019.

[46] R. T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *Journal on Optimization Theory and Applications*, 12:555–562, 1973.

[47] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285, 1974.