Comput Optim Appl (2018) 69:51–75
https://doi.org/10.1007/s10589-017-9937-2

Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points

E. G. Birgin[1] · G. Haeser[2,3] · A. Ramos[4]

Received: 7 December 2016 / Published online: 16 August 2017

Abstract Augmented Lagrangian methods with convergence to second-order stationary points in which any constraint can be penalized or carried out to the subproblems are considered in this work. The resolution of each subproblem can be done by any numerical algorithm able to return approximate second-order stationary points. The developed global convergence theory is stronger than the ones known for current algorithms with convergence to second-order points in the sense that, besides the flexibility introduced by the general lower-level approach, it includes a loose requirement for the resolution of subproblems. The proposed approach relies on a weak constraint qualification, that allows Lagrange multipliers to be unbounded at the solution. It is also shown that second-order resolution of subproblems increases the chances of finding a feasible point, in the sense that limit points are second-order stationary for the problem

This work has been partially supported by FAPESP (Grants 2013/03447-6, 2013/05475-7, 2013/07375-0, 2016/01860-1, and 2016/02092-8) and CNPq (Grants 481992/2013-8, 309517/2014-1, and 303264/2015-2).

✉ E. G. Birgin
egbirgin@ime.usp.br

G. Haeser
ghaeser@ime.usp.br

A. Ramos
albertoramos@ufpr.br

1 Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, São Paulo, SP 05508-090, Brazil

2 Department of Applied Mathematics, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, São Paulo, SP 05508-090, Brazil

3 Department of Management Science and Engineering, Stanford University, Stanford, CA, USA

4 Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil

of minimizing the squared infeasibility. The applicability of the proposed method is illustrated in numerical examples with ball-constrained subproblems.

Keywords Augmented Lagrangian methods · Nonlinear programming · Second-order stationary points · Algorithms

1 Introduction

In general, algorithms for nonlinear optimization are iterative. A simplified model is solved at each iteration with the goal of building a sequence of iterates converging to a solution of the original problem. Usually, the more strict one solves the subproblems, the better is the quality of the solution found. If one requires subproblems to be solved up to first-order stationarity, under reasonable assumptions, the final solution is also first-order stationary. In this paper, our goal is to increase the quality of the final solution to second-order stationarity by obtaining second-order stationary points of the subproblems. This is particularly relevant in some machine and statistical learning problems [48], where a second-order stationary point is a provably good substitute for an actual solution, or in problems with multiple non-optimal first-order stationary points.

We will consider a large class of Augmented Lagrangian algorithms, in the sense that we do not specify which constraints are penalized (upper-level constraints) and which constraints are kept in the subproblems (lower-level constraints). Theoretically, the division into lower- and upper-level constraints can be done arbitrarily [24]. However, many practical motivations exist that drive the separation of the constraint into those two sets. The Augmented Lagrangian iterates are feasible, or approximately feasible, with respect to the lower-level constraints; while the upper-level constraints are, in general, satisfied only in the limit. It may be the case, depending on the application, that the objective function is meaningless when the lower-level constraints are violated; so they must be satisfied at all iterations. In [26], lower-level constraints are called non-relaxable; while upper-level constraints are called relaxable constraints. An extreme case is when the objective function is not defined outside the lower-level constraints. In practice, the choice of lower- and upper-level constraints is strongly based on the possibility of solving subproblems with lower-level constraints. It may also depend on convenience if, for instance, in a particular application, some well established and tuned solver exists and an additional constraint the solver is not capable to deal with must be considered. In this situation, it would be natural to adopt an Augmented Lagrangian algorithm penalizing the new constraint and using the well established solver for solving the subproblems. Two particular examples of this situation follow. The first one is the important problem of minimizing a matrix function subject to an orthogonality constraint. Fast tailored algorithms exist for this problem; but they can not be easily generalized when additional constraints are present. (See [37] and references therein for details.) A second example is bi-level optimization (see, for example, [35]), where one wants to minimize an objective function under the constraint that some variables are a solution to an optimization problem parameterized by the other variables. When additional standard constraints are present, they are natural candidates to be upper-level constraints; since there are efficient algorithms for bi-level optimization problems without additional constraints.

First-order Augmented Lagrangian methods in which the lower-level constraints are bound constraints can be found in [5,32], while second-order ones can be found in [6]. More recently, an Augmented Lagrangian method that keeps equality constraints at the lower-level set and penalizes the remaining constraints (including the bound constraints) was presented in [19]. This motivated us to further develop the global convergence theory of Augmented Lagrangian methods with arbitrary constraints at the lower-level set.

In first-order optimization methods, the Karush–Kuhn–Tucker (KKT) conditions are the standard first-order stationarity concept. In the second-order case, the situation is not straightforward. In theory, many second-order optimality conditions can be derived, with or without the presence of constraint qualifications and with distinct properties. Conditions may rely on (i) generalized multipliers associated with the objective function (also called Fritz John multipliers [29, Theorem 3.50]), (ii) the whole set of Lagrange multipliers, (iii) only one Lagrange multiplier [3,15,57], (iv) the critical cone, or (v) the critical subspace (weak critical cone). See [16] for a discussion and more details on second-order optimality conditions.

The second-order condition that will be considered in the present work is the weak second-order condition (WSOC). It states that the Hessian of the Lagrangian at a KKT point is positive semidefinite on the subspace orthogonal to the gradients of active constraints. Several reasons lead us to consider this second-order condition. Usually, algorithms build a single approximation of a Lagrange multiplier and not the whole set of Lagrange multipliers. Hence, conditions depending on the whole set of Lagrange multipliers may be hard to check in practice. We also do not consider Fritz John multipliers, since algorithms for solving subproblems usually treat objective function and constraints in different manners, in a way that a multiplier associated with the objective function is not present. Finally, checking the positive semidefiniteness of a matrix over a subspace is a relatively easy task. In contrast, checking the positive semidefiniteness of a matrix on a pointed cone is an NP-hard problem [56]. These considerations make WSOC a natural second-order condition whose approximate satisfaction may be considered as stopping criterion for practical algorithms. Other than that, up to the authors knowledge, no algorithm with global convergence to points satisfying a stronger second-order stationarity concept is known. Moreover, the discussion in [40], where a very simple algorithm is shown to converge to a point that violates a strong optimality condition, suggests that such algorithms may not exist.

Second-order methods for constrained optimization have been devised in [6,14, 38,39,55]. In particular, the methods proposed in [6,55] are Augmented Lagrangian methods. The method proposed in the present work differs from those methods in at least two aspects—it has the flexibility introduced by the general lower-level approach and its global convergence theory relies on weak assumptions. In particular, previous results are based on the Mangasarian-Fromowitz constraint qualification, which is equivalent to the boundedness of Lagrange multipliers; while in this work global convergence is guaranteed even when the solution has unbounded Lagrange multipliers. When the algorithm converges to an infeasible point, it is shown that the limit point is second-order stationary for the problem of minimizing an infeasibility measure, which also enhances previous results. Since subproblems are not solved exactly, the presented analysis relies on a perturbation of the second-order optimality conditions defined in [9].

The rest of this work is organized as follows. Section 2 describes the Augmented Lagrangian framework and gives some basic definitions. Section 3 presents the global convergence results. A discussion and numerical illustrative examples are given in Sect. 4. Section 5 gives some conclusions.

Notation We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space. By $\mathbb{R}^n_+$ we denote the set of elements of $\mathbb{R}^n$ for which each component is non-negative. The notation $\max\{0, a\}$ represents the maximum between $a \in \mathbb{R}$ and 0. If $a \in \mathbb{R}^n$, $\max\{0, a\}$ is to be taken componentwise. The symbol $\|\cdot\|$ stands for the Euclidean norm and $\|\cdot\|_\infty$ stands for the supremum norm. Given a set-valued mapping $\mathcal{F} : \mathbb{R}^s \rightrightarrows \mathbb{R}^\ell$, the *sequential Painlevé-Kuratowski outer limit* of $\mathcal{F}(z)$ as $z \to z^*$ is denoted by

$$\limsup_{z \to z^*} \mathcal{F}(z) = \{w^* \in \mathbb{R}^\ell : \exists\ (z^k, w^k) \to (z^*, w^*) \quad \text{with } w^k \in \mathcal{F}(z^k)\}.$$

We say that $\mathcal{F}$ is *outer semicontinuous* at z^* if $\limsup_{z \to z^*} \mathcal{F}(z) \subseteq \mathcal{F}(z^*)$.

2 Augmented Lagrangian framework

Consider the optimization problem

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega_L \cap \Omega_U, \tag{OP}$$

where $\Omega_U = \{x \in \mathbb{R}^n \mid H(x) = 0, G(x) \le 0\}$ and $\Omega_L = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \le 0\}$. We assume that $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $H : \mathbb{R}^n \to \mathbb{R}^{\bar{m}}$ and $G : \mathbb{R}^n \to \mathbb{R}^{\bar{p}}$ are twice continuously differentiable.

Constraints in Ω_U are the so-called upper-level constraints and they will be penalized with the use of the Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian function [43,59,61] defined by

$$x \mapsto L(x, \bar{v}, \bar{u}; \rho) = f(x) + \frac{\rho}{2}\left(\left\|H(x) + \frac{\bar{v}}{\rho}\right\|^2 + \left\|\max\left\{0, G(x) + \frac{\bar{u}}{\rho}\right\}\right\|^2\right).$$

The main algorithm consists in a sequence of approximate minimizations of the Augmented Lagrangian $L(x, \bar{v}, \bar{u}; \rho)$ over the set Ω_L followed by an update of the multipliers $\bar{v} \in \mathbb{R}^{\bar{m}}$ and $\bar{u} \in \mathbb{R}^{\bar{p}}_+$ and the penalty parameter $\rho > 0$. This means that the penalty parameter ρ and the Lagrange multipliers approximations $\bar{v}$ and $\bar{u}$ drive the feasibility with respect to Ω_U; while the feasibility of the lower-level constraints Ω_L is kept in each iteration. Each approximate minimization of $L(x, \bar{v}, \bar{u}; \rho)$ is called an outer iteration. For updating the multipliers approximations, we adopt the classical first-order update [17,43,59] but, following [4–6], we consider a safeguarded multiplier estimate laying in a compact box. To update the penalty parameter ρ, we consider a slight generalization of the update in [4–6] (see also [24]). We continue to describe the Augmented Lagrangian framework.

Algorithm 2.1. Augmented Lagrangian framework [4,5,24]

Let $v_{\min} < v_{\max}$, $u_{\max} > 0$, $\gamma > 1$, $\tau \in [0, 1)$, $\bar{v}^1 \in [v_{\min}, v_{\max}]^{\bar{m}}$, and $\bar{u}^1 \in [0, u_{\max}]^{\bar{p}}$ be given. Initialize $k \leftarrow 1$.

Step 1 Find and approximate solution x^k to

$$\text{Minimize } L(x, \bar{v}^k, \bar{u}^k; \rho_k) \text{ subject to } x \in \Omega_L. \tag{1}$$

Step 2 Define

$$v^k = \bar{v}^k + \rho_k H(x^k) \text{ and } u^k = \max\{0, \bar{u}^k + \rho_k G(x^k)\}$$

and compute $\bar{v}^{k+1} \in [v_{\min}, v_{\max}]^{\bar{m}}$ and $\bar{u}^{k+1} \in [0, u_{\max}]^{\bar{p}}$ (typically, as the projections of v^k and u^k onto the corresponding safeguarded boxes).

Step 3 Set

$$V^k = \max\left\{G(x^k), -\bar{u}^k/\rho_k\right\}.$$

If $k = 1$ or

$$\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|H(x^{k-1})\|_\infty \|V^{k-1}\|_\infty\}$$

then choose ρ_{k+1} satisfying $\rho_{k+1} \geq \rho_k$. Otherwise, choose ρ_{k+1} such that $\rho_{k+1} \geq \gamma\rho_k$.

Step 4 Set $k \leftarrow k + 1$ and go to Step 1.

The measure V^k is a joint measure of feasibility (with respect to the inequality constraints) and complementarity. Thus, $\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\}$ measures the failure of the feasibility and the complementarity at the iterate x^k. Clearly, $\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} = 0$ if, and only if, $H(x^k) = 0$, $G(x^k) \leq 0$, and $(\bar{u}^k)^T G(x^k) = 0$. In order to control the growth of the penalty parameter ρ_k, it is increased when the joint measure of feasibility and complementarity is not sufficiently reduced. Safeguarded Lagrange multipliers approximations are considered in order to obtain convergence to a global minimizer when subproblems are globally solved. See [5] or [24] for further details on this framework. Note that, despite the safeguarded strategy, the Lagrange multipliers approximations v^k and u^k may still be unbounded. See [45] for a comparison of Augmented Lagrangian methods with and without safeguarded Lagrange multipliers.

In order to formulate precisely our general Augmented Lagrangian framework, we must clarify what is meant by an approximate solution to subproblem (1). For a first-order Augmented Lagrangian method, this means that x^k is an ε_k-KKT point, for a given sequence of tolerances $\varepsilon_k \to 0^+$, as defined below. Let us consider subproblem

(1) with a general objective function, that is, let us consider the general optimization problem with a continuously differentiable objective function $F : \mathbb{R}^n \to \mathbb{R}$ given by

$$\text{Minimize } F(x) \text{ subject to } h(x) = 0 \text{ and } g(x) \leq 0. \tag{GOP}$$

Definition 2.1 (*ε-KKT*) Given $\varepsilon \geq 0$, we say that $x \in \mathbb{R}^n$ is an ε-KKT point for the problem (GOP) if there exist $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that

$$\|h(x)\| \leq \varepsilon, \ \|\max\{0, g(x)\}\| \leq \varepsilon, \tag{2}$$

$$\left\| \nabla F(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{i=1}^{p} \mu_i \nabla g_i(x) \right\| \leq \varepsilon, \tag{3}$$

$$\mu_i = 0 \text{ whenever } g_i(x) < -\varepsilon, \quad i = 1, \ldots, p. \tag{4}$$

Note that ε-KKT for $\varepsilon = 0$ is equivalent to the KKT conditions and that (2), (3), and (4) are perturbed measures of feasibility, optimality, and complementarity, respectively. An important property of ε-KKT points is that, given a local minimizer x^* of (GOP), for every $\varepsilon > 0$, there exists a point $x_\varepsilon \in \mathbb{R}^n$ with $\|x^* - x_\varepsilon\| < \varepsilon$ such that x_ε is an ε-KKT point. That is, even though x^* may not be a KKT point, due, for instance, to the failure of a constraint qualification, it can be arbitrarily approximated by ε-KKT points. This motivates the following definition. See [7].

Definition 2.2 (*Approximate-KKT* [7]) We say that $x^* \in \mathbb{R}^n$ is an Approximate-KKT (AKKT) point for the problem (GOP) if there exist sequences $\varepsilon_k \to 0^+$ and $x^k \to x^*$ such that x^k is ε_k-KKT for all k. Equivalently, a feasible point x^* is an AKKT point if there exist sequences $x^k \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, and $\{\mu^k\} \subset \mathbb{R}^p_+$ such that

$$\nabla F(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in \mathcal{A}(x^*)} \mu_i^k \nabla g_i(x^k) \to 0,$$

where $\mathcal{A}(x^*) = \{i \in \{1, \ldots, p\} \mid g_i(x^*) = 0\}$.

Theorem 2.1 ([7]) *If x^* is a local minimizer then x^* is an AKKT point.*

Thus, the AKKT condition is a true optimality condition independently of the validity of any constraint qualification. Theorem 2.1 justifies the stopping criterion based on the KKT condition which is used in several optimization methods for solving constrained optimization problems, i.e. stopping when finding an ε-KKT point for a given $\varepsilon > 0$ small enough. A local minimizer may not be a KKT point but we can always find a point sufficiently close to the local minimizer satisfying approximately the KKT condition, which makes it reasonable to accept it as possible solution.

The global convergence theory of the first-order version of Algorithm 2.1 is based on the optimality condition AKKT as described below.

Theorem 2.2 ([5]) *Let $\varepsilon_k \to 0^+$ be a sequence of tolerances. Assume that $\{x^k\}$ is a sequence generated by Algorithm* 2.1 *and that, at Step* 1, *x^k is computed as an*

ε_k-KKT point for problem (1)*, that is, for every k,* (2–4) *hold with $x = x^k$, $\varepsilon = \varepsilon_k$, and $F(x) = L(x, \bar{v}^k, \bar{u}^k; \rho_k)$. Then, every limit point x^* of $\{x^k\}$ is an AKKT point of the infeasibility optimization problem*

$$\textit{Minimize } \|H(x)\|^2 + \|\max\{0, G(x)\}\|^2 \textit{ subject to } x \in \Omega_L. \tag{IOP}$$

It is worth noting that the sequence of tolerances $\{\varepsilon_k\}$ may be defined in an external or an internal way (depending on the feasibility of the current iterate). For further details in adaptive ways for choosing the subproblems tolerances ε_k, see, for example, [21,25,36].

Theorem 2.3 ([5]) *Assume that $\{x^k\}$ is a sequence generated by Algorithm* 2.1 *as in Theorem* 2.2*. Then, every feasible limit point $x^* \in \Omega_L \cap \Omega_U$ of $\{x^k\}$ is an AKKT point of the original optimization problem* (OP)*.*

In order to measure the strength of the optimality condition AKKT, it can be compared with classical optimality conditions based on constraint qualifications. This can be achieved with the following result.

Definition 2.3 (*CCP* [11]) We say that the Cone Continuity Property (CCP) with respect to the constraints of problem (GOP) holds at a feasible point x^* when the set-valued mapping $x \mapsto K(x)$ is outer semicontinuous at x^*, that is, $\limsup_{x \to x^*} K(x) \subseteq K(x^*)$, where

$$K(x) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{i \in \mathcal{A}(x^*)} \mu_i \nabla g_i(x) \mid \mu \in \mathbb{R}^p_+, \lambda \in \mathbb{R}^m \right\}.$$

Theorem 2.4 ([11]) *The Cone Continuity Property holds at x^* with respect to the constraints of problem* (GOP) *if, and only if, for every objective function $F(x)$ in* (GOP) *such that x^* is AKKT, x^* is also a KKT point.*

In particular, assuming x^* satisfies CCP with respect to the constraints in Ω_L and with respect to $\Omega_L \cap \Omega_U$, allows us to rewrite Theorem 2.2 and 2.3 obtaining true KKT points and, therefore, these results can be compared with other results that rely on a constraint qualification. Note that Theorems 2.1 and 2.4 imply that CCP is a constraint qualification and that CCP is weaker than most of the other constraint qualifications used to analyze global convergence of algorithms. It is strictly weaker than CPG [10], CRSC [10], (R)CPLD [8,60], (R)CRCQ [44,54], MFCQ [49], and LICQ [18]. For details, see [11] and the references therein.

The next section is devoted to developing an analogous second-order global convergence theory for the Augmented Lagrangian method with general lower-level constraints when subproblems are approximately solved up to second-order.

3 A second-order Augmented Lagrangian

In this section, we consider that, at Step 1 of Algorithm 2.1, the approximate solution x^k is obtained taking into account second-order information. To be precise, let us consider again subproblem (1) with a general objective function given in (GOP).

For $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$, let us consider the Lagrangian function $x \mapsto \ell(x, \lambda, \mu)$ associated with (GOP) given by

$$\ell(x, \lambda, \mu) = F(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{p} \mu_i g_i(x).$$

We recall that WSOC holds at a feasible point x^* for the problem (GOP) if there exists Lagrange multipliers $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ with $\mu_i = 0$ whenever $g_i(x^*) < 0$ such that

$$\nabla F(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i \in \mathcal{A}(x^*)} \mu_i \nabla g_i(x^*) = 0 \tag{5}$$

and, in addition, for each d in the critical subspace

$$S = \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^{\mathrm{T}} d = 0, \quad i = 1, \ldots, m, \text{ and } \nabla g_i(x^*)^{\mathrm{T}} d = 0, \quad i \in \mathcal{A}(x^*)\},$$

we have that

$$d^T \left(\nabla^2 F(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*) + \sum_{i \in \mathcal{A}(x^*)} \mu_i \nabla^2 g_i(x^*) \right) d \geq 0. \tag{6}$$

Similar to the KKT condition, WSOC holds at a local minimizer only under some constraint qualification. However, not all constraint qualifications serve this purpose. For instance, WSOC holds under LICQ but not under MFCQ. See [16] for more details on WSOC. Based on this notion of second-order stationarity, a second order version of AKKT (called AKKT2) was presented in [9]. Its definition is given below.

Definition 3.1 (*AKKT2* [9]) We say that the feasible point $x^* \in \mathbb{R}^n$ fulfills AKKT2 with respect to (GOP) if there exist sequences $x^k \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, $\{\theta^k\} \subset \mathbb{R}^m_+$, $\{\eta^k\} \subset \mathbb{R}^p_+$, and $\delta_k \to 0^+$, with $\mu_i^k = 0$ and $\eta_i^k = 0$ whenever $g_i(x^*) < 0$, such that

$$\nabla \ell(x^k, \lambda^k, \mu^k) = \nabla F(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in \mathcal{A}(x^*)} \mu_i^k \nabla g_i(x^k) \to 0 \tag{7}$$

and

$$\nabla^2 \ell(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \theta_i^k \nabla h(x^k) \nabla h(x^k)^T + \sum_{i \in \mathcal{A}(x^*)} \eta_i^k \nabla g(x^k) \nabla g(x^k)^T + \delta_k \mathcal{I} \tag{8}$$

is positive semidefinite for all k, where $\mathcal{I}$ is the identity matrix.

In [9], it was proved that AKKT2 is an optimality condition, that is, when x^* is a local minimizer, there must exist sequences given by Definition 3.1, even if WSOC fails at x^*. Thus, AKKT2 justifies a practical stopping criterion based on the approximate fulfillment of WSOC. In this section, we prove global convergence results for Algorithm 2.1 based on the optimality condition AKKT2.

In order to measure the strength of AKKT2, we recall the definition of the second-order cone-continuity property (CCP2) from [9].

Definition 3.2 (*CCP2* [9]) We say that CCP2 holds at a feasible point x^* with respect to the problem (GOP) when the set-valued mapping $x \mapsto K_2(x)$ is outer semicontinuous at x^*, i.e. $\limsup_{x \to x^*} K_2(x) \subseteq K_2(x^*)$, where $K_2(x)$ is the cone given by

$$\bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+ \\ \mu_i=0,\ i\notin\mathcal{A}(x^*)}} \left\{ \begin{array}{c} \text{all } (w, W) \text{ with } w\in\mathbb{R}^n \text{ and } W \text{ a symmetric } n\times n \text{ matrix such that} \\ w+\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{i\in\mathcal{A}(x^*)} \mu_i \nabla g_i(x) = 0, \\ d^T\left(W+\sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{i\in\mathcal{A}(x^*)} \mu_i \nabla^2 g_i(x)\right) d \geq 0 \quad \text{for all } d\in S(x,x^*) \end{array} \right\} \tag{9}$$

and

$$S(x,x^*) = \left\{ d\in\mathbb{R}^n \mid \nabla h_i(x)^T d = 0, \quad i=1,\ldots,m, \text{ and } \nabla g_i(x)^T d = 0, \quad i\in\mathcal{A}(x^*) \right\} \tag{10}$$

is the perturbed critical subspace around the point x^*.

The cone $K_2(x^*)$ allows us to rewrite WSOC in a geometrical way. In fact, a feasible point x^* satisfies WSOC if, and only if, there exists $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ with $\mu_i = 0$ whenever $g_i(x^*) < 0$, such that $\nabla \ell(x^*, \lambda, \mu) = 0$ and $\nabla^2 \ell(x^*, \lambda, \mu)$ is positive semidefinite over the critical subspace $S(x^*, x^*) = S$, which can be equivalently stated, shortly, as $(\nabla F(x^*), \nabla^2 F(x^*)) \in K_2(x^*)$. The following theorem states precisely the strength of AKKT2.

Theorem 3.1 ([9]) *CCP2 holds at x^* with respect to the constraints of problem* (GOP) *if, and only if, for all objective functions $F(x)$ in* (GOP) *such that x^* is AKKT2, x^* also satisfies WSOC.*

In [9], it was proved that CCP2 is strictly weaker than the joint condition MFCQ+WCR, CRCQ [44], and RCRCQ [54] and that it can be used in the global convergence analysis of the second-order algorithms defined in [6,12,39]. In particular, in view of Theorem 3.1, global convergence results under AKKT2 can be restated in terms of the usual WSOC under CCP2. This means that even when MFCQ fails, and hence the solution does not have bounded multipliers, a classical global convergence result is still available.

We turn now to the issue of developing an alternative definition of AKKT2 that relies on ε perturbations of the second-order stationarity notion WSOC and that can be checked without the knowledge of the limit point x^*. This would allow us to provide a tolerance for the solvability of the subproblems at Step 1 of Algorithm 2.1 that guarantees that limit points of a sequence generated by the algorithm satisfy AKKT2.

Lemma 3.1 ([17,34]) *Let P be a symmetric $n \times n$ matrix and $a_1, \dots, a_r \in \mathbb{R}^n$. Define the linear subspace $\mathcal{C} = \{d \in \mathbb{R}^n \mid a_i^T d = 0, i = 1, \dots, r\}$. Suppose that $v^T P v > 0$ for all $v \in \mathcal{C}$. Then, there exist positive scalars c_i, $i = 1, \dots, r$, such that $P + \sum_{i=1}^r c_i a_i a_i^T$ is positive definite.*

Now, we provide an equivalent definition of AKKT2 based on ε perturbations. This equivalence can be used as stopping criterion for constrained optimization algorithms.

Theorem 3.2 *A point x^* satisfies AKKT2 for problem (GOP) if, and only if, there are sequences $x^k \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, and $\varepsilon_k \to 0^+$ such that*

$$\|h(x^k)\| \le \varepsilon_k, \ \|\max\{0, g(x^k)\}\| \le \varepsilon_k, \tag{11}$$

$$\|\nabla \ell(x^k, \lambda^k, \mu^k)\| = \left\| \nabla F(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(x^k) \right\| \le \varepsilon_k, \tag{12}$$

$$\mu_i^k = 0 \text{ whenever } g_i(x^k) < -\varepsilon_k, \tag{13}$$

and

$$d^T \nabla^2 \ell(x^k, \lambda^k, \mu^k) d \ge -\varepsilon_k \|d\|^2, \tag{14}$$

for all $d \in S_k$, where S_k is the subspace given by

$$S_k = \left\{ d \in \mathbb{R}^n \ \middle| \begin{array}{l} \nabla h_i(x^k)^T d = 0, i = 1, \dots, m, \text{ and} \\ \nabla g_i(x^k)^T d = 0, i = 1, \dots, p \text{ such that } g_i(x^k) \ge -\varepsilon_k \end{array} \right\}. \tag{15}$$

Proof Let x^* be an AKKT2 point and let $\{x^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{\theta^k\}$, $\{\eta^k\}$, and $\{\delta_k\}$ be the sequences given by Definition 3.1. Define

$$\varepsilon_k = \max\left\{ \delta^k, \|h(x^k)\|, \|\max\{0, g(x^k)\}\|, \|\nabla \ell(x^k, \lambda^k, \mu^k)\|, \{-g_i(x^k), i \in \mathcal{A}(x^*)\} \right\}.$$

Clearly, we have that $\varepsilon_k \to 0^+$ and (11) and (12) hold. (13) also holds. In fact, if for some index i_0 we have $g_{i_0}(x^k) < -\varepsilon_k$, or, equivalently, $-g_{i_0}(x^k) > \varepsilon_k \ge \max_{\{i \in \mathcal{A}(x^*)\}}\{-g_i(x^k)\}$, then, $i_0 \notin \mathcal{A}(x^*)$. Thus, $\mu_{i_0}^k = 0$ by the definition of the sequence $\{\mu^k\}$. Let us note that for k large enough, $S_k = S(x^k, x^*)$. This comes from the fact that $g_i(x^*) = 0$ if, and only if, $g_i(x^k) \ge -\varepsilon_k$ for k large enough. The implication $g_i(x^k) < -\varepsilon_k \Rightarrow g_i(x^*) < 0$ has already been proved; while the converse is a simple consequence of the continuity of g_i and $\varepsilon_k \to 0^+$. Note that $S(x^k, x^*) \subseteq S_k$ holds eventually for any sequence $\varepsilon_k \to 0^+$; while the reciprocal inclusion is a consequence of this particular definition of ε_k. For k large enough, let $d \in S_k = S(x^k, x^*)$. Using (8) and the definition of $S(x^k, x^*)$ we get $d^T \nabla^2 \ell(x^k, \lambda^k, \mu^k) d \ge -\delta_k \|d\|^2 \ge -\varepsilon_k \|d\|^2$ and (14) holds.

Now let us assume that x^* is such that there are sequences $x^k \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, and $\varepsilon_k \to 0^+$ such that (11–14) hold. The continuity of h and g together with (11) imply that x^* is feasible. Also, if $g_i(x^*) < 0$, eventually, $g_i(x^k) < -\varepsilon_k$ for k large enough. Hence, (12) and (13) imply (7). From (14), for k large enough, the matrix $P_k = \nabla^2 \ell(x^k, \lambda^k, \mu^k) + (\varepsilon_k + \frac{1}{k})\mathcal{I}$ is positive definite on the subspace $S_k \supseteq S(x^k, x^*)$. Hence, defining $\delta_k = \varepsilon_k + \frac{1}{k} \to 0^+$, Lemma 3.1 with matrix P_k and the subspace $S(x^k, x^*)$ implies the existence of θ_i^k, $i = 1, \ldots, m$, and η_i^k, $i \in \mathcal{A}(x^*)$, such that (8) holds. □

Note that a stronger equivalence can be proved without using the same tolerance ε_k in all equations (11–14), as long as all the considered tolerances go to zero. Hence, Theorem 3.2 suggests the following stopping criterion for methods solving constrained optimization problems with convergence to second-order stationary points.

Given tolerances $\varepsilon_{\text{feas}} > 0$, $\varepsilon_{\text{opt}} > 0$, $\varepsilon_{\text{compl}} > 0$, and $\varepsilon_{\text{curv}} > 0$ for feasibility, optimality, complementarity, and curvature, respectively, accept x^k as an approximation to a solution when

$$\|h(x^k)\| \leq \varepsilon_{\text{feas}}, \ \|\max\{0, g(x^k)\}\| \leq \varepsilon_{\text{feas}}, \tag{16}$$

$$\|\nabla \ell(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_{\text{opt}}, \tag{17}$$

$$\mu_i^k = 0 \text{ whenever } g_i(x^k) < -\varepsilon_{\text{compl}}, \tag{18}$$

and

$$d^T \nabla^2 \ell(x^k, \lambda^k, \mu^k) d \geq -\varepsilon_{\text{curv}} \|d\|^2, \tag{19}$$

for all $d \in \mathbb{R}^n$ with $\nabla h_i(x^k)^T d = 0, i = 1, \ldots, m$, and $\nabla g_i(x^k)^T d = 0$ whenever $g_i(x^k) \geq -\varepsilon_{\text{compl}}$. A verification of (19) can be done under mild computational cost, for instance, computing the smallest eigenvalue of the reduced matrix $\nabla^2 \ell(x^k, \lambda^k, \mu^k)$ in the corresponding subspace (see, for example, [47] and the references therein). Similarly to the first-order case, when (11–14) hold for some $\lambda^k \in \mathbb{R}^m$ and $\mu^k \in \mathbb{R}^p_+$, we say that x^k is ε_k-KKT2 for problem (GOP). Note that when x is ε-KKT2 with $\varepsilon = 0$, x satisfies WSOC.

In order to prove our second-order global convergence results based on AKKT2, a natural way is to assume that subproblem (1) is approximately solved up to second-order, that is, x^k is ε_k-KKT2 with $\varepsilon_k \to 0^+$. This is reasonable since it can be expected that algorithms with convergence to second-order stationary points will generate AKKT2 limit points, see [9]. But there is a minor issue about the possible lack of second-order differentiability of the objective function $F(x) = L(x, \bar{v}^k, \bar{u}^k; \rho_k)$ of subproblems (1). The problem here is that $\max\{0, G_i(x) + \bar{u}_i^k/\rho_k\}^2$ may not be two times differentiable at points x where $G_i(x) + \bar{u}_i^k/\rho_k = 0$. To precisely formulate our results, let us consider the general subproblem (GOP) where the objective function $F(x)$ is of the PHR form, that is,

$$F(x) = f_0(x) + \frac{1}{2} \sum_{i=1}^{r} \max\{0, f_i(x)\}^2,$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable for $i = 0, 1, \ldots, r$. Following [6, Prop. 1], we can see that the matrix

$$\nabla^2 F_\varepsilon(x) = \nabla^2 f_0(x) + \sum_{i=1}^{r} \max\{0, f_i(x)\}\nabla^2 f_i(x) + \sum_{f_i(x)\geq -\varepsilon} \nabla f_i(x)\nabla f_i(x)^T, \quad \varepsilon \geq 0, \tag{20}$$

can play the role of a Hessian matrix of F in its Taylor expansion. In particular, second-order optimality conditions for the minimization of $F(x)$ can be formulated as if $F(x)$ were twice continuously differentiable using $\nabla^2 F_\varepsilon(x)$ instead of the Hessian of $F(x)$. The optimality condition depends on $\varepsilon \geq 0$ and is more stringent the closer ε is to zero. When F is twice continuously differentiable, $\nabla^2 F = \nabla^2 F_0$, that is, the Hessian of F at x coincides with the expression in (20) with $\varepsilon = 0$.

We will say that x^* is an AKKT2 point for problem (GOP) with the PHR objective function $F(x)$ when there are appropriate sequences such that (11–14) hold with $\nabla^2 F(x^k)$ replaced by $\nabla^2 F_{\varepsilon_k}(x^k)$, or, what can be shown to be equivalent, when (7–8) hold with $\nabla^2 F(x^k)$ replaced by

$$\nabla^2 f_0(x^k) + \sum_{i=1}^{r} \max\{0, f_i(x^k)\}\nabla^2 f_i(x^k) + \sum_{\{i \mid f_i(x^*)\geq 0\}} \nabla f_i(x^k)\nabla f_i(x^k)^T.$$

It is easy to see from the proof of Theorem 3.3 in [9] that AKKT2 for PHR functions in the way that we have defined is an optimality condition. Also, Theorem 3.1 still holds for PHR functions, where $\nabla^2 F(x)$ in the definition of WSOC is replaced by $\nabla^2 F_0(x)$.

In the following theorem we prove that limit points of the second-order Augmented Lagrangian algorithm are second-order stationary points for the problem of minimizing the squared infeasibility measure, which is a PHR function. This extends previous results where only first-order conditions were obtained in [6] for the particular case of a compact box at the lower-level.

Theorem 3.3 *Assume that $\varepsilon_k \to 0^+$ is a sequence of given tolerances and that $\{x^k\}$ is a sequence generated by Algorithm* 2.1 *in such a way that, at Step 1, x^k is ε_k-KKT2 for subproblem* (1)*, i.e.* (11–14) *hold with $F(x) = L(x, \bar{v}^k, \bar{u}^k; \rho_k)$ and $\nabla^2 F(x^k) = \nabla^2 F_{\varepsilon_k}(x^k)$. Then, every limit point x^* of $\{x^k\}$ is an AKKT2 point for the infeasibility optimization problem*

$$\textit{Minimize } \|H(x)\|^2 + \|\max\{0, G(x)\}\|^2 \textit{ subject to } x \in \Omega_L. \tag{IOP}$$

Proof Assume without loss of generality that $x^k \to x^*$. From (11), we get $x^* \in \Omega_L$. Now, we will examine two cases depending on whether the sequence $\{\rho_k\}$ is bounded or not.

If $\{\rho_k\}$ is bounded, Step 3 of Algorithm 2.1 ensures that $\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} \to 0$. Hence $H(x^k) \to 0$ and $V^k = \max\{G(x^k), -\bar{u}^k/\rho_k\} \to 0$. Thus, $H(x^*) = 0$, $G(x^*) \leq 0$, and x^* is a global minimizer of (IOP), in particular, an AKKT2 point.

Now, we consider the case in which the sequence $\{\rho_k\}$ is unbounded. By some calculations and using the notation $\nabla^2 L(x^k, \bar{v}^k, \bar{u}^k; \rho_k) = \nabla^2 L_{\varepsilon_k}(x^k, \bar{v}^k, \bar{u}^k; \rho_k)$, we get

$$\begin{aligned}\nabla L(x^k, \bar{v}^k, \bar{u}^k; \rho_k) = \nabla f(x^k) &+ \sum_{i=1}^{\bar{m}} (\bar{v}_i^k + \rho_k H_i(x^k)) \nabla H_i(x^k) \\ &+ \sum_{i=1}^{\bar{p}} \max\{0, \bar{u}_i^k + \rho_k G_i(x^k)\} \nabla G_i(x^k)\end{aligned}$$

and

$$\begin{aligned}\nabla^2 L(x^k, \bar{v}^k, \bar{u}^k; \rho_k) = \nabla^2 f(x^k) &+ \sum_{i=1}^{\bar{m}} (\bar{v}_i^k + \rho_k H_i(x^k)) \nabla^2 H_i(x^k) \\ &+ \sum_{i=1}^{\bar{p}} \max\{0, \bar{u}_i^k + \rho_k G_i(x^k)\} \nabla^2 G_i(x^k) \\ &+ \sum_{i=1}^{\bar{m}} \rho_k \nabla H_i(x^k) \nabla H_i(x^k)^T \\ &+ \sum_{\{i \mid \bar{u}_i^k + \rho_k G_i(x^k) \geq -\varepsilon_k\}} \rho_k \nabla G_i(x^k) \nabla G_i(x^k)^T.\end{aligned}$$

Substituting $\nabla L(x^k, \bar{v}^k, \bar{u}^k; \rho_k)$ and $\nabla^2 L(x^k, \bar{v}^k, \bar{u}^k; \rho_k)$ in (12) and (14), we get

$$\|L_k\| \leq \varepsilon_k \tag{21}$$

and

$$d^T M_k d \geq -\varepsilon_k \|d\|^2 \text{ for all } d \in S_k, \tag{22}$$

where

$$\begin{aligned} S_k = \{&d \in \mathbb{R}^n \mid \nabla h_i(x^k)^T d = 0, \\ &i = 1, \ldots, m, \text{ and } \nabla g_i(x^k)^T d = 0 \text{ whenever } g_i(x^k) \geq -\varepsilon_k\}, \\ L_k = \nabla f(x^k) &+ \sum_{i=1}^{\bar{m}} (\bar{v}_i^k + \rho_k H_i(x^k)) \nabla H_i(x^k) \\ &+ \sum_{i=1}^{\bar{p}} \max\{0, \bar{u}_i^k + \rho_k G_i(x^k)\} \nabla G_i(x^k) \\ &+ \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla g_i(x^k),\end{aligned}$$

$$
\begin{aligned}
M_k = {} & \nabla^2 f(x^k) + \sum_{i=1}^{\bar{m}} (\bar{v}_i^k + \rho_k H_i(x^k)) \nabla^2 H_i(x^k) \\
& + \sum_{i=1}^{\bar{p}} \max\{0, \bar{u}_i^k + \rho_k G_i(x^k)\} \nabla^2 G_i(x^k) + \sum_{i=1}^{\bar{m}} \rho_k \nabla H_i(x^k) \nabla H_i(x^k)^T \\
& + \sum_{\{i \mid \bar{u}_i^k + \rho_k G_i(x^k) \geq -\varepsilon_k\}} \rho_k \nabla G_i(x^k) \nabla G_i(x^k)^T \\
& + \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla^2 g_i(x^k).
\end{aligned}
$$

Now, we proceed by dividing (21) and (22) by ρ_k. Noting that $\rho_k^{-1}(\nabla f(x^k), \nabla^2 f(x^k), \bar{v}^k, \bar{u}^k, \varepsilon_k)$ vanishes in the limit, given that $\{\rho_k\}$ is unbounded and $\{\bar{v}^k\}$ and $\{\bar{u}^k\}$ are bounded by Step 2 of Algorithm 2.1, we get

$$
\left\| \sum_{i=1}^{\bar{m}} H_i(x^k) \nabla H_i(x^k) + \sum_{i=1}^{\bar{p}} \max\{0, G_i(x^k)\} \nabla G_i(x^k) + \sum_{i=1}^{m} \frac{\lambda_i^k}{\rho_k} \nabla h_i(x^k) + \sum_{i=1}^{p} \frac{\mu_i^k}{\rho_k} \nabla g_i(x^k) \right\| \leq \varepsilon_k'
$$

and

$$
\begin{aligned}
d^T \Big(& \textstyle\sum_{i=1}^{\bar{m}} H_i(x^k) \nabla^2 H_i(x^k) + \sum_{i=1}^{\bar{p}} \max\{0, G_i(x^k)\} \nabla^2 G_i(x^k) \\
& + \textstyle\sum_{i=1}^{\bar{m}} \nabla H_i(x^k) \nabla H_i(x^k)^T + \sum_{\{i \mid \bar{u}_i^k + \rho_k G_i(x^k) \geq -\varepsilon_k\}} \nabla G_i(x^k) \nabla G_i(x^k)^T \\
& + \textstyle\sum_{i=1}^{m} \frac{\lambda_i^k}{\rho_k} \nabla^2 h_i(x^k) + \sum_{i=1}^{p} \frac{\mu_i^k}{\rho_k} \nabla^2 g_i(x^k) \Big) d \geq -\varepsilon_k'' \|d\|^2 \quad \text{for all } d \in S_k,
\end{aligned} \tag{23}
$$

for some $\varepsilon_k' \to 0^+$ and $\varepsilon_k'' \to 0^+$. Since, on the one hand, if $G_i(x^*) < 0$ then we have that $\max\{0, G_i(x^k)\} = 0$ and that $\bar{u}_i^k + \rho_k G_i(x^k) < -\varepsilon_k$ for sufficiently large k, and, on the other hand, $\nabla G_i(x^k) \nabla G_i(x^k)^T$ is positive semidefinite, we can replace

$$
\sum_{i=1}^{\bar{p}} \max\{0, G_i(x^k)\} \nabla^2 G_i(x^k) \quad \text{and} \quad \sum_{\{i \mid \bar{u}_i^k + \rho_k G_i(x^k) \geq -\varepsilon_k\}} \nabla G_i(x^k) \nabla G_i(x^k)^T
$$

by

$$
\sum_{\{i \mid G_i(x^*) \geq 0\}} \max\{0, G_i(x^k)\} \nabla^2 G_i(x^k) \quad \text{and} \quad \sum_{\{i \mid G_i(x^*) \geq 0\}} \nabla G_i(x^k) \nabla G_i(x^k)^T
$$

in (23), respectively, yielding that x^* is an AKKT2 point of (IOP). □

Now, it remains to prove the relationship between the sequence of iterates given by Algorithm 2.1 and the original problem (OP). The next theorem makes the precise statement.

Theorem 3.4 *Let $\{x^k\}$ be a sequence generated by Algorithm 2.1 as in Theorem 3.3. Then, every feasible limit point $x^* \in \Omega_L \cap \Omega_U$ of $\{x^k\}$ is an AKKT2 point of the original problem (OP).*

Proof As usual, there is no loss of generality if we assume that $\{x^k\}$ converges to $x^* \in \Omega_L \cap \Omega_U$. Following the proof of Theorem 3.3, we get that (21) and (22) hold, where L_k and M_k are defined as in the proof of Theorem 3.3.

Set $v^k = \bar{v}^k + \rho_k H(x^k)$ and $u^k = \max\{0, \bar{u}^k + \rho_k G(x^k)\} \geq 0$. Now, we will prove that $u_i^k = 0$ whenever $G_i(x^*) < 0$. We have two cases depending on whether the sequence of penalty parameters $\{\rho_k\}$ is bounded or not.

If $\{\rho_k\}$ is bounded then Step 3 of the algorithm implies that $V^k = \max\{G(x^k), -\bar{u}^k/\rho_k\} \to 0$. Hence, if $G_{i_0}(x^*) < 0$, for some i_0, we have that $\bar{u}_{i_0}^k \to 0$ and $\bar{u}_{i_0}^k + \rho_k G_{i_0}(x^k) < -\varepsilon_k$ for k large enough. Thus, for every i such that $G_i(x^*) < 0$, we get $u_i^k = 0$. If $\{\rho_k\}$ is unbounded, we also have $\bar{u}_i^k + \rho_k G_i(x^k) < -\varepsilon_k$ for k large enough whenever $G_i(x^*) < 0$. From both cases, we get that if $G_i(x^*) < 0$ then $u_i^k = 0$ for k large enough.

Since for k large enough, $u_i^k = 0$, $\bar{u}_i^k + \rho_k G_i(x^k) < -\varepsilon_k$ whenever $G_i(x^*) < 0$, and $\mu_i^k = 0$ if $g_i(x^*) < 0$, we can write L_k and M_k as follows:

$$
\begin{aligned}
L_k &= \nabla f(x^k) + \sum_{i=1}^{\bar{m}} v_i^k \nabla H_i(x^k) + \sum_{\{i \mid G_i(x^*)=0\}} u_i^k \nabla G_i(x^k) \\
&\quad + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{\{i \mid g_i(x^*)=0\}} \mu_i^k \nabla g_i(x^k), \\
M_k &= \nabla^2 f(x^k) + \sum_{i=1}^{\bar{m}} v_i^k \nabla^2 H_i(x^k) + \sum_{\{i \mid G_i(x^*)=0\}} u_i^k \nabla^2 G_i(x^k) \\
&\quad + \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{\{i \mid g_i(x^*)=0\}} \mu_i^k \nabla^2 g_i(x^k) \\
&\quad + \sum_{i=1}^{\bar{m}} \rho_k \nabla H_i(x^k) \nabla H_i(x^k)^T + \sum_{\{i \mid G_i(x^*)=0\}} \rho_k \nabla G_i(x^k) \nabla G_i(x^k)^T.
\end{aligned}
$$

Furthermore, by (21) and (22), we obtain that $L_k \to 0$ and $d^T M_k d \geq -\varepsilon_k \|d\|$ for all $d \in S_k$, where S_k is the subspace given in (15). To end the proof, consider the subspace W_k defined as

$$
W_k = \left\{ d \in \mathbb{R}^n \,\middle|\, \begin{array}{llll} \nabla H_i(x^k)^T d = 0, & i = 1, \ldots, \bar{m}, & \nabla G_i(x^k)^T d = 0 & \text{whenever } G_i(x^*) = 0, \\ \nabla h_i(x^k)^T d = 0, & i = 1, \ldots, m, & \text{and } \nabla g_i(x^k)^T d = 0, & \text{whenever } g_i(x^*) = 0 \end{array} \right\}.
$$

Clearly, $g_i(x^k) \geq -\varepsilon_k$ for all k sufficiently large imply $g_i(x^*) = 0$. Thus, $W_k \subseteq S_k$. Then, the matrix $P_k = M_k + (\varepsilon_k + \frac{1}{k})\mathcal{I}$ is positive definite on W_k. We get the desired result applying Lemma 3.1 for the matrix P_k and the subspace W_k. □

As a consequence of Theorems 3.1, 3.3, and 3.4 we obtain the following result.

Corollary 3.1 *Let $\{x^k\}$ be a sequence generated by Algorithm* 2.1 *in such a way that, at Step* 1*, we require that x^k is ε_k-KKT2 for $\varepsilon_k \to 0^+$. Let x^* be a limit point of $\{x^k\}$.*

1. *If x^* satisfies CCP2 for the infeasibility optimization problem (IOP) then x^* satisfies WSOC with respect to (IOP).*
2. *If x^* is feasible and satisfies CCP2 for the original problem (OP) then x^* satisfies WSOC with respect to (OP).*

Corollary 3.1 allows us to obtain a second-order stationarity result even if in a solution the set of Lagrange multipliers is unbounded, that is, MFCQ fails. Furthermore, since RCRCQ implies CCP2, the theory developed above guarantees convergence to second-order stationary points when all the constraints are given by affine functions.

In order to get a stronger global convergence theory, we can impose a slightly stronger condition to accept x^k as an approximate solution to the kth subproblem, as well as an additional smoothness assumption on the upper-level constraints. In [13], it has been suggested to replace (4) in the definition of ε-KKT by the stronger condition given by

$$|\lambda_i h_i(x)| < \varepsilon, \quad i = 1, \ldots, m, \quad \text{and } |\mu_i g_i(x)| < \varepsilon, \quad i = 1, \ldots, p. \tag{24}$$

When (2), (3), and (24) hold for some $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$, x is called an ε-CAKKT point, where the letter C stands for "complementarity". A limit of ε-CAKKT points with $\varepsilon \to 0^+$ is called a CAKKT point. In [13], it was proved that CAKKT is an optimality condition and that feasible limit points of a general lower-level Augmented Lagrangian are CAKKT points of the original problem, given that an additional smoothness assumption on the upper-level constraints is satisfied. In order to obtain those results, it was also assumed that each x^k is an ε_k-KKT point and that the lower-level constraints are simple in the sense that Lagrange multipliers approximations associated with them can be assumed to be bounded.

Let us consider now the generalization of CAKKT to second-order given in [42]. We say that $x \in \mathbb{R}^n$ is an ε-CAKKT2 point if, for some $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$, $\theta \in \mathbb{R}^m_+$, and $\eta \in \mathbb{R}^p_+$, conditions (2), (3), and (24) hold together with

$$\nabla^2 \ell(x, \lambda, \mu) + \sum_{i=1}^{m} \theta_i \nabla h(x) \nabla h(x)^T + \sum_{i=1}^{p} \eta_i \nabla g(x) \nabla g(x)^T + \varepsilon \mathcal{I} \tag{25}$$

being positive semidefinite and

$$|\theta_i h_i(x)^2| < \varepsilon, \quad i = 1, \ldots, m, \quad \text{and } |\eta_i g_i(x)^2| < \varepsilon, \quad i = 1, \ldots, p. \tag{26}$$

A limit of ε-CAKKT2 points with $\varepsilon \to 0^+$ is called a CAKKT2 point. In [42] it was proved that CAKKT2 is an optimality condition strictly stronger than CAKKT and AKKT2 together. Let us present now the global convergence result of Algorithm 2.1 when subproblems are solved obtaining, at each iteration, ε_k-CAKKT or ε_k-CAKKT2 points with $\varepsilon_k \to 0^+$, under the same additional smoothness property on the upper-level constraints needed in the first-order case.

Theorem 3.5 *Assume that $\varepsilon_k \to 0^+$ is a sequence of given tolerances and that $\{x^k\}$ is a sequence generated by Algorithm 2.1 where, at Step 1, x^k is ε_k-CAKKT2 (or ε_k-CAKKT) for subproblem (1), with $F(x) = L(x, \bar{v}^k, \bar{u}^k; \rho_k)$ and $\nabla^2 F(x^k) = \nabla^2 F_{\varepsilon_k}(x^k)$. Then, any limit point x^* of $\{x^k\}$ is a CAKKT2 point (or CAKKT point, respectively) for the infeasibility optimization problem (IOP). If, in addition, we assume (the generalized Łojasiewicz inequality that says) that there exist $\delta > 0$ and a function $\varphi : B(x^*, \delta) \to \mathbb{R}$ with $\varphi(x) \to 0$ as $x \to x^*$ such that for all $x \in B(x^*, \delta)$,*

$$|\Phi(x) - \Phi(x^*)| \le \varphi(x)\|\nabla\Phi(x)\|,$$

where $\Phi(x) = \|H(x)\|^2 + \|\max\{0, G(x)\}\|^2$, then, when x^* is feasible for the original problem (OP), x^* is a CAKKT2 point (or CAKKT point, respectively) of the original problem (OP).

Proof For the lower-level constraints, the additional property follows easily from the additional requirements (24) and (26) for the subproblems. When x^* is feasible, under the generalized Łojasiewicz inequality, we can follow the proof of [13, Theorem 5.1] to see that $\rho_k H_i(x^k)^2 \to 0, i = 1, \dots, \bar{m}$, $\rho_k G_i(x^k)^2 \to 0, i = 1, \dots, \bar{p}$, $v_i^k H_i(x^k) \to 0, i = 1, \dots, \bar{m}$, and $u_i^k G_i(x^k) \to 0, i = 1, \dots, \bar{p}$. Now, defining $\theta_i^k = \rho_k, i = 1, \dots, \bar{m}$, and $\eta_i^k = \rho_k$ if $\bar{u}_i^k + \rho_k G_i(x^k) \ge -\varepsilon_k$, $\eta_i^k = 0$ otherwise, we can see that $\theta_i^k H_i(x^k)^2 \to 0, i = 1, \dots, \bar{m}$, and $\eta_i^k G_i(x^k)^2 \to 0, i = 1, \dots, \bar{p}$, as in the proof of [42, Theorem 3.1]. □

We end this section with a detailed comparison of our results with the global convergence results presented for the Augmented Lagrangian method introduced in [6]. The following improvements apply: (a) a weaker constraint qualification is employed, (b) lower-level constraints can be arbitrary, (c) subproblems can be solved slightly more loosely, and (d) a second-order stationarity result also holds for the infeasibility problem.

The requirements employed in this work for an approximate solution to the subproblems are slightly weaker than the ones required in [6, Theorem 2], where the particular case of a compact box in the lower-level is considered. The condition described in [6] for solving the subproblems corresponds to replacing the set S_k by the larger one $\{d \in \mathbb{R}^n \mid \nabla h_i(x^k)^T d = 0, i = 1, \dots, m,$ and $\nabla g_i(x^k)^T d = 0$ whenever $g_i(x^k) \ge 0\}$, that requires more from the subproblem solver. However, it should be noted that one could modify the definition of AKKT2 by replacing the range of the last sum in (8) from $\mathcal{A}(x^*) = \{i \in \{1, \dots, p\} \mid g_i(x^*) = 0\}$ to the smaller index set $\{i \in \{1, \dots, p\} \mid g_i(x^k) \ge 0\}$, arriving to a stronger optimality condition. The condition considered in [6] to accept x^k as an approximate subproblem's solution is precisely the ε_k-perturbation of this modified AKKT2 condition (analogous to the ε_k-perturbation of the AKKT2 considered in Theorem 3.2). The fact that this is an optimality condition follows from the proof of [9, Theorem 3.3]. The point is that it is not the case that feasible limit points of the Augmented Lagrangian method fulfill this stronger optimality condition, independently of the requirements made to consider x^k an approximate solution to the kth subproblem. This makes the formulation being introduced in the present work more natural, in the sense that the (weaker) criterion used for the resolution of the subproblems coincides with the optimality condition present in the global convergence theorem.

Similarly, with respect to the definition of AKKT2 for objective functions of the PHR form, we note that although the optimality condition is stronger replacing $\nabla^2 F(x^k)$ by $\nabla^2 F_0(x^k)$, it is required less from an approximate solution to a subproblems by using $\nabla^2 F_{\varepsilon_k}(x^k)$, which makes it easier to the subproblem solver to stop and it is also compatible with the AKKT2 definition for PHR functions.

4 Discussion and illustrative numerical examples

In many practical constrained optimization problems the constraints can be divided in two classes: the ones that may be satisfied only at the end of the optimization process and the ones that should be satisfied at every stage of it. The reason for that is that the fulfillment of the first type of constraints is not mandatory, for many different reasons; while the fulfillment of the second ones cannot be avoided, for example, because they involve definitions of variables. Intermediate results in which the first constraints are not fully satisfied may be useful in practice just because the second ones are. In those cases it is not sensible to deal with the constraints in identical ways which means that, in the terminology of Augmented Lagrangian methods, the mandatory constraints should be permanently included in the so called lower-level set, or classified as subproblem constraints. See [24]. Therefore, at each subproblem of the Augmented Lagrangian method one should solve a subproblem in which the only constraints are those of the second class.

The family of Augmented Lagrangian methods introduced in the present work possesses two desirable features at the same time—convergence to second-order stationary points and general constraints in the so-called lower-level set. In the present section, we illustrate with numerical examples the possible advantages of both features. However, it should be noted that the implementation of this kind of method is limited by the existence of a method able to tackle the subproblems. Candidates should exhibit convergence to second-order stationary points and should be able to deal with the lower-level constraints by preserving feasibility of the iterates. While, on the one hand, feasible second-order methods for bound-constrained minimization exist (see, for example, [6]); on the other hand, feasible second-order methods with arbitrary constraints are rare. In the present section, we focus on situations in which the lower-level constraints are given by Euclidean balls and, therefore, subproblems can be solved by the trust-region method introduced in [50,51].

4.1 Avoiding convergence to global maximizers

Regarding the convergence to second-order stationary points, it is not easy to exhibit numerical experiments in which a method with convergence to second-order stationary points finds better quality solutions (feasible points with smaller objective function value) than those found by a method with convergence to first-order stationary points. A recent numerical experiment presented in [23], considering *all* the 87 unconstrained problems of the CUTEst collection [41] with available second-order derivatives, illustrates this fact. In this experiment it was observed that every time a first-order stopping criterion was satisfied, a second-order stopping criterion was satisfied as well at the same time. The consequence of this was that the first- and second-order methods being compared obtained equivalent results in all the problems in which both satisfied a stopping criterion related to success (independently of being a first- or a second-order stopping criterion). *Ad hoc* problems in which a first-order method converges to a global maximizer were presented in [6] and [23], but this behavior is hardly observed in massive numerical experiments. A similar example follows.

Consider the problem

$$\text{Minimize } x \text{ subject to } x + y^2 = 0 \quad \text{and} \quad x^2 + y^2 \leq 1.$$

It has a global maximizer at $x = y = 0$ and global minimizers at $x = (1 - \sqrt{5})/2$ and $y = \pm\sqrt{(\sqrt{5} - 1)/2}$. Algencan [5] (see also [24]) is a well-established implementation of an Augmented Lagrangian method that keeps the bound constraints as lower-level constraints (and penalizes any other type of constraint) and solves the subproblems with a variety of methods for bound-constrained minimizations, all of them with convergence to first-order stationary points. When Algencan is applied to the problem above starting from $(2, 0)^T$, due to the problem's symmetry with respect to the x-axis, it converges to the global maximizer $(x, y)^T = (0, 0)^T$ with $\lambda_1 = -1$ and $\mu_1 = 0$. (A similar behaviour would be expected from any other first-order method.) Consider now the possibility of applying an Augmented Lagrangian method that keeps $x^2 + y^2 \leq 1$ as a lower-level constraint and finds second-order stationary points of the subproblems. The subproblem at the kth iteration is of the form

$$\text{Minimize } x + \frac{\rho_k}{2}\left(x + y^2\right)^2 \text{ subject to } x^2 + y^2 \leq 1,$$

where, with the simple purpose of simplifying the presentation, we consider $\bar{v}_1^k = 0$ for all k (assuming that parameters $\bar{v}_{\min}$ and $\bar{v}_{\max}$ are such that $0 \in [\bar{v}_{\min}, \bar{v}_{\max}]$). Figure 1 shows a sequence of approximate second-order stationary points for the subproblems, for increasing values of ρ_k, that converges to one of the global minimizers. In theory, a global minimizer would also be obtained by penalizing both constraints, if one were able to find second-order stationary points for the subproblems. However, in practice, if a "small" penalty parameter is used for the first subproblem, the greediness phenomenon described below impairs convergence in this case.

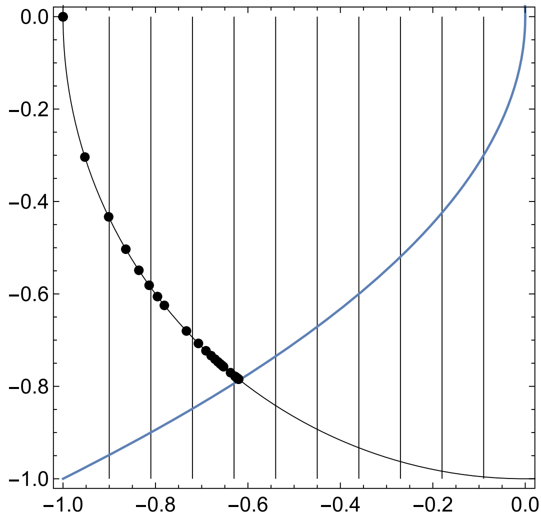

Fig. 1 Sequence $\{x^k\}$ of second-order stationary points of the kth subproblem for increasing values of ρ_k. The whole sequence satisfies the lower-level constraint $x^2 + y^2 \leq 1$. The infeasible point $x^1 \approx (-1, 0)^T$ (with respect to the upper-level constraint $x + y^2 = 0$) corresponds to a small value of $\rho_1 = 0.1$; while the final iterate $\approx ((1 - \sqrt{5})/2, -((\sqrt{5} - 1)/2)^{1/2})^T$ corresponds to $\rho_k = 100$ and it satisfies the penalized constraint with tolerance 10^{-8}

4.2 Avoiding the greediness phenomenon

Penalty and Augmented Lagrangian methods may be affected by the behavior of the objective function outside the feasible region. If the objective function takes very low values at infeasible points, iterates of the subproblem solver may be attracted by undesirable minimizers or they may even diverge, especially at the first outer iterations, and overall convergence may fail to occur. This phenomenon was named "greediness" in [31], where a Proximal Augmented Lagrangian method was proposed as an alternative to overcome it. In [18, pp. 418–419], it was suggested as a remedy that an external penalty function with exponent bigger than 2 should be used; while, in [20], an outer trust-region approach for the same purpose was introduced. None of these alternatives is necessary if the greediness phenomenon is caused by the penalization of a constraint that can be kept at the lower-level set.

Consider the problem

$$\text{Minimize } x^5 \text{ subject to } x^2 \leq 1 \text{ and } x \leq 0.$$

If the bound constraint is kept at the lower-level set and the ball constraint is penalized, as it is the case in Algencan, subproblems are of the form

$$\text{Minimize } x^5 + \frac{\rho_k}{2}\left(x^2 - 1\right)^2 \text{ subject to } x \leq 0.$$

Clearly, subproblems are unbounded from below for any value of ρ_k and, therefore, the Augmented Lagrangians' theory does not help to predict the practical behavior of the method in this case. Even adding an artificial lower bound of the form $-10^{20} \leq x$ may not help to avoid the greediness phenomenon. If $x = -10^{20}$ is obtained as a solution to the first subproblem and, as it is usual in practice, it is used as initial guess for solving the next subproblem, many outer iterations and increases of the penalty parameter will be needed to obtain a solution to the original problem. Moreover, it is well-known that large penalty parameters promote ill-conditioned subproblems, numerical issues, and, eventually, practical failure of the inner solver to satisfy the subproblems' stopping criterion. This is what happens with Algencan. On the other hand, if the ball constraint can be kept as a lower-level constraint, subproblems are of the form

$$\text{Minimize } x^5 + \frac{\rho_k}{2} \max\{0, x\}^2 \text{ subject to } x^2 \leq 1$$

and, in this particular case, convergence to the global minimizer $x = -1$ is expected to occur at the first outer iteration.

4.3 Dealing with functions that are undefined outside a given domain

Consider an objective function $f : \mathbb{R}^n \to \mathbb{R}$ that should be minimized within $\Omega_1 \cap \Omega_2$ and such that f or its derivatives are not well-defined outside Ω_1. In this case, it appears to be no other choice other than setting the lower-level set $\Omega_L = \Omega_1$ and the upper-level set $\Omega_U = \Omega_2$; i.e. subproblems must consist on minimizing the Augmented Lagrangian that penalizes Ω_2 subject to Ω_1. Three simple examples follow.

Example 1 In [27], a location problem was introduced. The problem consists in finding locations $c_0, c_1, \ldots, c_N \in \mathbb{R}^2$ such that $\sum_{i=1}^{N} \|c_0 - c_i\|_2$ is minimized subject to $c_i \in C_i \cap \Omega_i$ for $i = 0, \ldots, N$, where $C_i = \{x \in R^2 \mid \|c_i - \bar{c}_i\| \leq r_i\}$ for given constants $\bar{c}_i \in \mathbb{R}^2$ and $r_i \in \mathbb{R}$ and it is assumed that $C_0 \cap C_i = \emptyset$ for all $i > 0$. The constraints Ω_i are some additional constraints that must be satisfied by each location c_i, $i = 0, \ldots, N$. If all constraints are penalized, subproblems are unconstrained and they are not differentiable if $c_0 = c_i$ for some $i > 0$. This inconvenience is avoided if constraints $c_i \in C_i$ for $i = 0, \ldots, N$ are kept as lower-level constraints.

Example 2 Similar examples occur when the potential energy function $\sum_{i \neq j} 1/\|p_i - p_j\|^2$ is minimized in order to obtain "well-distributed" points p_i within a given region (see, for example, [58], where this function was considered as an ingredient for solving packing problems). It may be the case that keeping some constraints in the lower-level set avoids the regions where the objective functions is not well-defined ($p_i = p_j$ for some $i \neq j$).

Example 3 There exists a huge variety of packing problem. They basically consist in placing items within a container without overlapping. One of the simplest packing problems consists in maximizing the radius r of N identical circles that must be packed within a given container. On the one hand, this simple geometrical problem has received the attention of several researchers in the last decades (see [28] and the references therein). On the other hand, this problem is an over-simplification of a packing problem that applies to building initial configurations for molecular dynamics simulations [52,53] (see also [24, p. 165]).

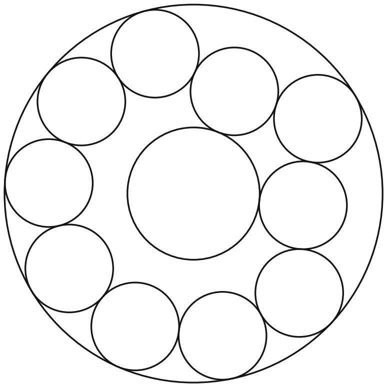

Fig. 2 $N = 10$ identical circles of maximizing radius $r \approx 0.4586$ packed within a *ring-shape* container given by an *inner circle* with radius $R_{\text{in}} = 0.7$ and an *outer circle* with radius $R_{\text{out}} = 2$

Consider the problem of packing identical circles within the "ring-shape" region displayed in Fig. 2. The problem can be modelled as

$$\begin{array}{ll} \text{Maximize } r & \\ \text{subject to} & (x_i - x_j)^2 + (y_i - y_j)^2 \geq (2r)^2 \quad \text{for } i = 1, \ldots, N, \quad j = i+1, \ldots, N, \\ & x_i^2 + y_i^2 \leq (R_{\text{out}} - r)^2 \quad \text{for } i = 1, \ldots, N, \\ & x_i^2 + y_i^2 \geq (R_{\text{in}} + r)^2 \quad \text{for } i = 1, \ldots, N. \end{array}$$

The first constraint avoids the overlapping between the circles being packed; while the two remaining constraints represent the fact that the items must be placed within the ring-shape container with an inner circle of radius R_{in} and an outer circle of radius R_{out}. Constraint $x_i^2 + y_i^2 \geq (R_{\text{in}} + r)^2$ may also be written as

$$\log[(R_{\text{out}} - r)^2 - (x_i^2 + y_i^2)] \leq c, \tag{27}$$

where $c = \log[(R_{\text{out}} - r)^2 - (R_{\text{in}} + r)^2]$. This equivalent constraint is not well-defined if $x_i^2 + y_i^2 \leq (R_{\text{out}} - r)^2$ does not hold and, therefore, the latter constraint should be kept as a lower level constraint. Of course, (27) is an unhappy modelling choice, but it illustrates a situation that may occur in more complex real applications.

5 Conclusions

Typical implementations of Augmented Lagrangian methods consider subproblems with simple constraints such as bound-constraints (as is the case of the well-known packages Algencan [5] and Lancelot [33]) or linear constraints (see, for example, [22]). However, in a recent work [19], the general lower-level theory of the first-order Augmented Lagrangian developed in [5] has been exploited to take a radical change on the classical approach, by penalizing the box constraints and keeping the equality constraints as subproblems' constraints. Similarly to the interior point approach, a Newtonian method applied to the KKT conditions of the subproblems has been implemented, with numerical results comparable to the classical approach employed by Algencan. This state of facts motivated us to further develop the second-order theory of Augmented Lagrangian methods considered in [6] including general lower-level constraints and using modern second-order optimality conditions [9,42] associated with weak constraint qualifications.

Several recently-developed affordable algorithms for unconstrained optimization [1,2,6,23,30,46] compute the least eigenvalue of the Hessian matrix at each iteration, and the corresponding eigenvector when necessary, in order to improve the complexity of finding a first-order stationary point, or to guarantee convergence to a second-order stationary point. Dealing with second-order information is not straightforward in general constrained optimization, but an analogous case would be when an algorithm computes directions of negative curvature in the kernel of the Jacobian of active lower-level constraints. In this case, one should expect that a second-order stopping criterion could be checked at a reasonable cost. This paper provided an Augmented Lagrangian framework to deal with constraints in those cases. Our future research will focus on developing an affordable second-order variant of the Newtonian Augmented Lagrangian with equality-constrained subproblems developed in [19], exploiting directions of negative curvature and the general theory developed in this paper.

References

1. Agarwal, N., Allen-Zhu, Z., Bullins, B., Hazan, E., Ma T.: Finding local minima for nonconvex optimization in linear time. arXiv: 1611.01146
2. Anandkumar, A., Ge, R.: Efficient approaches for escaping higher order saddle points in non-convex optimization. arXiv: 1602.05908v1
3. Andreani, R., Behling, R., Haeser, G., Silva, P.J.S.: On second order optimality conditions for nonlinear optimization. Optim. Methods Softw. **32**, 22–38 (2017)
4. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. **112**, 5–32 (2008)
5. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On Augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2008)
6. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. Comput. Optim. Appl. **45**, 209–236 (2010)
7. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)
8. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of optimization algorithms. IMA J. Numer. Anal. (2017). doi:10.1093/imanum/drw064
9. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**, 255–273 (2012)
10. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**, 1109–1135 (2012)
11. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**, 96–110 (2016)
12. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. Optimization **56**, 529–542 (2007)
13. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**, 3533–3554 (2010)
14. Avelino, C.P., Moguerza, J.M., Olivares, A., Prieto, F.J.: Combining and scaling descent and negative curvature directions. Math. Program. **128**, 285–319 (2011)

15. Baccari, A., Trad, A.: On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. SIAM J. Optim. **15**, 394–408 (2005)
16. Behling, R., Haeser, G., Ramos, A., Viana, D.S.: On a conjecture in second-order optimality conditions. Available at Optimization Online
17. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Athena Scientific, Belmont (1996)
18. Bertsekas, D.P.: Nonlinear programming, 2d edn. Athenas Scientific, Belmont (1999)
19. Birgin, E.G., Bueno, L.F., Martínez, J.M.: Sequential equality-constrained optimization for nonlinear programming. Comput. Optim. Appl. **65**, 699–721 (2016)
20. Birgin, E.G., Castelani, E.V., Martinez, A.L.M., Martínez, J.M.: Outer trust-region method for constrained optimization. J. Optim. Theory Appl. **150**, 142–155 (2011)
21. Birgin, E.G., Fernandez, D., Martínez, J.M.: On the boundedness of penalty parameters in an Augmented Lagrangian method with lower level constraints. Optim. Methods Softw. **27**, 1001–1024 (2012)
22. Birgin, E.G., Floudas, C.A., Martínez, J.M.: Global minimization using an Augmented Lagrangian method with variable lower-level constraints. Math. Program. **125**, 139–162 (2010)
23. Birgin, E.G., Martínez, J.M.: The use of quadratic regularization with a cubic descent condition for unconstrained optimization. SIAM J. Optim. **27**, 1049–1074 (2017)
24. Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization, Fundamentals of Algorithms, vol. 10. Society for Industrial and Applied Mathematics, Philadelphia (2014). doi:10.1137/1.9781611973365
25. Birgin, E.G., Martínez, J.M., Prudente, L.F.: Augmented Lagrangians with possible infeasibility and finite termination for global nonlinear programming. J. Glob. Optim. **58**, 207–242 (2014)
26. Birgin, E.G., Martínez, J.M., Prudente, L.F.: Optimality properties of an Augmented Lagrangian method on infeasible problems. Comput. Optim. Appl. **60**, 609–631 (2015)
27. Birgin, E.G., Martínez, J.M., Raydan, M.: Algorithm 813: SPG–software for convex-constrained optimization. ACM Trans. Math. Softw. **27**, 340–349 (2001)
28. Birgin, E.G., Martínez, J.M., Ronconi, D.P.: Optimizing the packing of cylinders into a rectangular container: a nonlinear approach. Eur. J. Oper. Res. **160**, 19–33 (2005)
29. Bonnans, J.F., Shapiro, A.: Pertubation Analysis of Optimization Problems. Springer, New York (2000)
30. Carmon, Y., Duchi, J.C., Hinder, O., Sidford, A.: Accelerated methods for nonconvex optimization. ArXiv: 1611.00756
31. Castelani, E.V., Martinez, A.L., Martínez, J.M., Svaiter, B.F.: Addressing the greediness phenomenon in nonlinear programming by means of proximal Augmented Lagrangians. Comput. Optim. Appl. **46**, 229–245 (2010)
32. Conn, A.R., Gould, N.I.M.: Toint, PhL: A globally convergent Augmented Lagrangian algorithm for optimization with general constraints and simple bounds. SIAM J. Numer. Anal. **28**, 545–572 (1991)
33. Conn, A.R., Gould, N.I.M.: Toint, PhL: Lancelot: A Fortran Package for Large-Scale Nonlinear Optimization (Release A). Springer, Berlin (1992)
34. Debreu, G.: Definite and semidefinite quadratic forms. Econometrica **20**, 295–300 (1952)
35. Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht (2002)
36. Eckstein, J., Silva, P.J.S.: A practical relative error criterion for Augmented Lagrangians. Math. Program. **141**, 319–348 (2013)
37. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**, 303–353 (1998)
38. Facchinei, F., Lucidi, S.: Convergence to second order stationary points in inequality constrained optimization. Math. Oper. Res. **23**, 746–766 (1998)
39. Gill, P.E., Kungurtsev, V., Robinson, D.P.: A stabilized SQP method: global convergence. IMA J. Numer. Anal. **37**(1), 407–443 (2017). doi:10.1093/imanum/drw004
40. Gould, N.I.M., Conn, A.R.: Toint, PhL: A note on the convergence of barrier algorithms for second-order necessary points. Math. Program. **85**, 433–438 (1998)
41. Gould, N.I.M., Orban, D.: Toint, PhL: CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. Comput. Optim. Appl. **60**, 545–557 (2014)
42. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Available at Optimization Online
43. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
44. Janin, R.: Direction derivative of the marginal function in nonlinear programming. Math. Program. Stud. **21**, 127–138 (1984)

45. Kanzow, C., Steck, D.: An Example Comparing the Standard and Modified Augmented Lagrangian Methods, Preprint, Institute of Mathematics, University of Würzburg, Würzburg, Feb 2017. http://www.mathematik.uni-wuerzburg.de/~kanzow/paper/ExampleALM.pdf
46. Karas, E.W., Santos, S.A., Svaiter, B.F.: Algebraic rules for quadratic regularization of Newton's method. Comput. Optim. Appl. **60**, 343–376 (2015)
47. Lehoucq, R.B., Sorensen, D.C., Yang, C.: ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods, Software, Environments and Tools, vol. 6. Society for Industrial and Applied Mathematics, Philadelphia (1998)
48. Liu, H., Yao, T., Li, R., Ye, Y.: Folded concave penalized sparse linear regression: sparsity, statistical performance and algorithmic theory for local solutions. Math. Program. (2017). doi:10.1007/s10107-017-1114-y
49. Mangasarian, O.L., Fromovitz, S.: The Fritz-John necessary conditions in presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
50. Martínez, J.M., Santos, S.A.: A trust-region strategy for minimization on arbitrary domains. Math. Program. **68**, 267–301 (1995)
51. Martínez, J.M., Santos, S.A.: New convergence results on an algorithm for norm constrained regularization and related problems. RAIRO-Oper. Res. **31**, 269–294 (1997)
52. Martínez, L., Andrade, R., Birgin, E.G., Martínez, J.M.: Packmol: a package for building initial configurations for molecular dynamics simulations. J. Comput. Chem. **30**, 2157–2164 (2009)
53. Martínez, L., Martínez, J.M.: Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. J. Comput. Chem. **24**, 819–825 (2003)
54. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**, 429–440 (2011)
55. Moguerza, J.M., Prieto, F.J.: An Augmented Lagrangian interior-point method using directions of negative curvature. Math. Program. **95**, 573–616 (2003)
56. Murty, K.G., Kabadi, S.N.: Some NP-complete problems in quadratic and nonlinear programming. Math. Program. **39**, 117–129 (1987)
57. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)
58. Nurmela, K.J., Östergåd, P.R.J.: Packing up to 50 equal circles in a square. Discrete Comput. Geom. **18**, 111–120 (1997)
59. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
60. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
61. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control Optim. **12**, 268–285 (1974)