ON ACHIEVING STRONG NECESSARY SECOND-ORDER PROPERTIES IN NONLINEAR PROGRAMMING*

Andreas Fischer, Gabriel Haeser and Thiago P. Silveira

September 29th, 2023.

Abstract: Second-order necessary or sufficient optimality conditions for nonlinear programming are usually defined by means of the positive (semi-)definiteness of a quadratic form, or a maximum of quadratic forms, over the critical cone. However, algorithms for finding such second-order stationary points are currently unknown. Typically, an algorithm with a second-order property approximates a primal-dual point such that the Hessian of the Lagrangian evaluated at the limit point is, under a constraint qualification, positive semidefinite over the lineality space of the critical cone. This is in general a proper subset of the critical cone, unless one assumes strict complementarity, which gives a weaker necessary optimality condition. In this paper, a new strong sequential optimality condition is suggested and analyzed. Based on this, we propose a penalty algorithm which, under certain conditions, is able to achieve second-order stationarity with positive semidefiniteness over the whole critical cone, which corresponds to a strong necessary optimality condition. Although the algorithm we propose is somewhat of a theoretical nature, our analysis provides a general framework for obtaining strong second-order stationarity.

Keywords: nonlinear optimization, second-order optimality conditions, constraint qualifications, global convergence
Mathematics Subject Classification: *49K05, 65K10, 90C26, 90C30*

---

1 Introduction

When proposing a derivative-based algorithm for smooth constrained optimization problems, one must have in mind efficiency and robustness. In terms of robustness, it is clear that one does not expect that a local minimizer will always be found. Thus, algorithms typically aim at finding points satisfying some first- or second-order *necessary* optimality condition. The Karush-Kuhn-Tucker conditions are usually the standard first-order stationarity notion employed. However, there are different notions of second-order stationarity.

Most notions of second-order stationarity are somewhat of a theoretical nature, since it is very difficult to incorporate them in a practical algorithm, at least not without impairing efficiency. Thus, most algorithms possessing a second-order global convergence theory only consider the simplest of these conditions, namely, one that does not make use of the full second-order information. More specifically, instead of ensuring positive semidefiniteness of the Hessian of the Lagrangian over the whole critical cone, this property is assured only in a subspace contained in the critical cone. This is done essentially because dealing with the whole critical cone is a computationally challenging task, see [17].

In this paper we consider nonlinear optimization problems in finite dimensions with equality and inequality constraints, where the problem functions are twice continuously differentiable and we aim at designing a general framework that is able to find a point satisfying a strong second-order necessary optimality condition, that is, considering the whole critical cone, under reasonable assumptions. This is done by means of a penalty algorithm that keeps the inequality constraints within the subproblems. However, our approach is somewhat theoretical as we do not propose an algorithm for solving the subproblems. This task remains a challenging open problem. Nevertheless, the analysis we conduct is non-standard and it consists of a first step towards the more general goal.

The paper starts in Section 2 with a review of different second-order necessary optimality conditions; we focus in particular on the results relying on assumptions on the rank of the gradients of constraints nearby the local minimizer, in particular, we consider the well known constant rank constraint qualification (CRCQ [15]).

*The research of the first author is supported by the Volkswagen Foundation under grant No. 97 775. The second author is funded by CNPq and the São Paulo Research Foundation grant 2018/24293-0, while the third author is funded by the São Paulo Research Foundation grants 2017/12187-9 and 2020/00130-5.

In Section 3 we present a gentle introduction to the topic of *sequential optimality conditions* [3], which is the main tool we employ to achieve our results. Based on this discussion, we introduce new strong second-order necessary approximate KKT conditions that consider the whole critical cone. Under a constant rank condition, we prove that all local minimizers of the optimization problem satisfy one of these approximate KKT conditions while the other condition is satisfied by all strict local minimizers.

In Section 4, we recall from [7, 12] that a standard barrier method and a second-order augmented Lagrangian method are not able to guarantee the strong second-order condition, even if a strict local minimizer of the subproblems is found at each iteration. Then, we show that this phenomenon will not occur (under a constant rank condition) if only equality constraints are penalized. That is, we propose our framework for designing an algorithm that will achieve the strong second-order condition under reasonable assumptions. The assumptions we employ rely on the constant rank of sets of gradients of constraints and the objective function, together with the extended Mangasarian-Fromovitz constraint qualification (MFCQ).

Our notation is rather standard. We just mention that $\|\cdot\|$ always denotes the Euclidean norm.

2 On Second-Order Conditions

Let us consider the nonlinear programming problem

$$\begin{array}{lll} \text{Minimize} & f(x), & \\ \text{subject to} & h_i(x) = 0, & i = 1, \dots, m, \\ & g_j(x) \leq 0, & j = 1, \dots, p, \end{array} \tag{NLP}$$

where the functions $f, h_i, g_j : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable. By Ω, we denote the feasible set of (NLP). Moreover, for any $x \in \mathbb{R}^n$, the set

$$A(x) := \{j \in \{1, \dots, p\} \mid g_j(x) = 0\}$$

contains the indices of inequality constraints that are active at x.

To formulate second-order conditions, let us first introduce the generalized Lagrangian $L_0 : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^p_+ \to \mathbb{R}$ by

$$L_0(x, r, \lambda, \mu) := rf(x) + h(x)^\top \lambda + g(x)^\top \mu,$$

whereas the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+ \to \mathbb{R}$ is given by

$$L(x, \lambda, \mu) := f(x) + h(x)^\top \lambda + g(x)^\top \mu.$$

Now, for any $x \in \Omega$, the set $\Lambda_0(x)$ of Fritz John multipliers and the set $\Lambda(x)$ of Lagrange multipliers are defined as

$$\Lambda_0(x) := \left\{0 \neq (r, \lambda, \mu) \in \mathbb{R}_+ \times \mathbb{R}^m \times \mathbb{R}^p_+ \mid \nabla_x L_0(x, r, \lambda, \mu) = 0, g(x)^\top \mu = 0\right\}$$

and

$$\Lambda(x) := \{(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+ \mid \nabla_x L(x, \lambda, \mu) = 0, g(x)^\top \mu = 0\},$$

respectively. For any $x \in \Omega$, we further need the critical cone

$$C(x) := \{d \in \mathbb{R}^n \mid \nabla f(x)^\top d \leq 0,\ \nabla h(x)^\top d = 0, \nabla g_j(x)^\top d \leq 0 \text{ for all } j \in A(x)\}. \tag{2.1}$$

The next theorem provides a pair of *no-gap* second-order optimality conditions. It can be derived from [10].

Theorem 2.1. *Let $\bar{x} \in \Omega$ be given. Then the following assertions are valid:*

a) If $\Lambda_0(\bar{x}) \neq \emptyset$ and

$$\sup_{(r,\lambda,\mu)\in\Lambda_0(\bar{x})} d^\top \nabla^2_{xx} L_0(\bar{x}, r, \lambda, \mu) d > 0 \quad \text{for all } d \in C(\bar{x}) \setminus \{0\},$$

then $\bar{x}$ is a strict local minimizer of (NLP).

b) If $\bar{x}$ is a local minimizer of (NLP), then $\Lambda_0(\bar{x}) \neq \emptyset$ and

$$\sup_{(r,\lambda,\mu)\in\Lambda_0(\bar{x})} d^\top \nabla^2_{xx} L(\bar{x}, r, \lambda, \mu) d \geq 0 \quad \text{for all } d \in C(\bar{x}).$$

Although several research is based on Fritz John multipliers, in this paper we are interested in Lagrange multipliers. As usual, any (x, λ, μ) is called a Karush-Kuhn-Tucker (KKT) point of (NLP) if $x \in \Omega$ and $(\lambda, \mu) \in \Lambda(x)$.

To avoid the distinction between Fritz John and Lagrange multipliers, one may assume the well-known Mangasarian-Fromovitz constraint qualification (MFCQ), which can be stated at $\bar{x} \in \Omega$ as saying that there is no Fritz John multiplier $(r, \lambda, \mu) \in \Lambda_0(\bar{x})$ with $r = 0$. Notice also that a Fritz John multiplier (r, λ, μ) with $r \neq 0$ provides a Lagrange multiplier $(\lambda/r, \mu/r)$. That is, when $r \neq 0$, one may without loss of generality consider $r = 1$. In this sense, under MFCQ, the notions of Fritz John and Lagrange multipliers coincide and Theorem 2.1 gives rise to the following standard second-order necessary optimality condition:

Proposition 2.2. *Let $\bar{x}$ be a local minimizer of* (NLP) *that satisfies MFCQ. Then $\Lambda(\bar{x}) \neq \emptyset$ and*

$$\sup_{(\lambda,\mu)\in\Lambda(\bar{x})} d^\top \nabla^2_{xx} L(\bar{x},\lambda,\mu) d \geq 0 \quad \text{for all } d \in C(\bar{x}). \tag{2.2}$$

Instead of MFCQ, we may use the constant rank constraint qualification (CRCQ) from [15]. This leads to the following second-order necessary optimality condition, which is the basis for our main focus in this paper.

Proposition 2.3 ([2])**.** *Let $\bar{x}$ be a local minimizer of* (NLP) *that satisfies CRCQ. Then $\Lambda(\bar{x}) \neq \emptyset$ and, for any $(\lambda,\mu) \in \Lambda(\bar{x})$, it holds that*

$$d^\top \nabla^2_{xx} L(\bar{x},\lambda,\mu) d \geq 0 \quad \text{for all } d \in C(\bar{x}). \tag{2.3}$$

Notice that under the stronger assumption that the linear independence constraint qualification (LICQ) holds at $\bar{x}$, Proposition 2.3 follows trivially from Proposition 2.2 since LICQ implies MFCQ and that $\Lambda(\bar{x})$ is a singleton. Obviously, condition (2.3) is a stronger necessary optimality condition than (2.2). Under CRCQ, this stronger condition (2.3) holds basically because CRCQ implies that the value of the second-order form $\langle d, \nabla^2_{xx} L(\bar{x},\lambda,\mu) d\rangle$ in (2.2) is, for any $d \in C(\bar{x})$, invariant to the choice of $(\lambda,\mu) \in \Lambda(\bar{x})$, see [11] and the extended version of [9].

Under CRCQ the necessary optimality condition (2.3) would be rather suitable for the algorithmic practice than condition (2.2). This means, given an algorithm that generates a primal-dual sequence $\{(x^k,\lambda^k,\mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$, one is interested in proving that a limit point $(\bar{x},\lambda,\mu) \in \Omega \times \mathbb{R}^m \times \mathbb{R}^p_+$ of this sequence is such that $(\lambda,\mu) \in \Lambda(\bar{x})$ and the second-order condition (2.3) is satisfied. However, no such algorithm has yet been presented. Algorithms with convergence to some kind of second-order point usually find limit points that satisfy a weaker version of (2.3) (see [12] and the references therein), where the critical cone $C(\bar{x})$ is replaced by its lineality space

$$S(\bar{x}) := \{d \in \mathbb{R}^n \mid \nabla f(\bar{x})^\top d = 0, \nabla h(\bar{x})^\top d = 0, \nabla g_j(\bar{x})^\top d = 0 \text{ for all } j \in A(\bar{x})\}. \tag{2.4}$$

Clearly, this necessary optimality condition is less interesting than the one presented in Proposition 2.3, since no associated sufficient optimality condition is known and $S(\bar{x}) \subseteq C(\bar{x})$. Also, one is essentially not able to exploit the structure of $S(\bar{x})$ in order to prove the result of Proposition 2.3 with $C(\bar{x})$ replaced by $S(\bar{x})$ under a condition weaker than CRCQ. An exception (but assuming MFCQ) is the following result. Its formulation makes use of the matrix $M(x) \in \mathbb{R}^{n\times(m+|A(\bar{x})|)}$ with $M(x) := \big(\nabla h(x), \nabla g_{A(\bar{x})}(x)\big)$.

Proposition 2.4 ([9, 13, 16])**.** *Let $\bar{x}$ be a local minimizer of* (NLP) *which satisfies MFCQ. If*

$$\operatorname{rank}(M(x)) \leq \operatorname{rank}(M(\bar{x})) + 1$$

for all x sufficiently close to $\bar{x}$, then there exists $(\lambda,\mu) \in \Lambda(\bar{x})$ with

$$d^\top \nabla^2_{xx} L(\bar{x},\lambda,\mu) d \geq 0 \quad \text{for all } d \in S(\bar{x}). \tag{2.5}$$

Of course, the less theoretical value of the second-order condition given by (2.5) is somehow compensated by its numerical tractability, see the discussion in the extended version of [9]. That is, many practical algorithms are able to exploit the linear space structure of $S(\bar{x})$ in order to achieve (2.5) in a reasonable manner. Our goal in this paper is to develop an algorithm whose limit points guarantee the stronger second-order necessary optimality condition (2.3).

3 A Strong Sequential Optimality Condition

The study of global convergence of algorithms under weak assumptions can be done with the aid of *sequential optimality conditions* [3]. Let us say that one is first able to show that an algorithm generates a sequence $\{x^k\}$ satisfying some mathematical proposition $\mathcal{P}(\{x^k\})$. Typically, this proposition is associated with a perturbation of a necessary optimality condition. The second step would be to prove that whenever $\bar{x}$ is a local minimizer, there exists a sequence $\{z^k\}$ with $z^k \to \bar{x}$ so that the proposition $\mathcal{P}(\{z^k\})$ is valid. When defining $\mathcal{P}(\cdot)$, of course, one is interested in as strong as possible necessary optimality conditions, however, one must consider the additional requirement that the mathematical proposition must also be satisfied by the sequence generated by the algorithm of interest. Both of these steps can usually be done without assuming that the problem satisfies a constraint qualification and may serve as an adequate enough global convergence theory. This strategy has been applied in several contexts, in particular when the problem has no clear standard optimality condition, or when one needs a consistent way of perturbing the standard optimality conditions, say, in order to conduct a complexity analysis [14, 18]. This avoids constraint qualifications at all; however, a final step of the analysis may be done using a constraint qualification for measuring the strength of the optimality condition: one should prove that when a feasible point $\bar{x}$ satisfies a constraint qualification and it can be approximated by some sequence $z^k \to \bar{x}$ so that the mathematical proposition $\mathcal{P}(\{z^k\})$ holds, then $\bar{x}$ satisfies a standard first- or second-order stationarity condition (say, that there exists some $(\lambda, \mu) \in \Lambda(\bar{x})$ that satisfies (2.5)).

For instance, when the problem has only equality constraints, an algorithm may generate a sequence $\{x^k\}$ that satisfies the mathematical proposition

$$\mathcal{P}(\{x^k\}) := \left[h(x^k) \to 0 \quad \text{and} \quad \nabla f(x^k) + \sum_{i=1}^{m} \nabla h_i(x^k)\lambda_i^k \to 0 \text{ for some sequence } \{\lambda^k\} \subset \mathbb{R}^m \right]$$

and one can prove that a local minimizer $\bar{x}$ may be approximated by a sequence $z^k \to \bar{x}$ of this type, that is, such that $\mathcal{P}(\{z^k\})$ is satisfied. Notice that this necessary optimality condition is strictly stronger than the usual Fritz John condition, which opens the path to considering constraint qualifications strictly weaker than LICQ (more generally, without assuming MFCQ, if inequality constraints are considered in this example).

The final step measuring the strength of the sequential optimality condition may consist of proving that when $\bar{x}$ satisfies some constraint qualification and there exists at least one sequence $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds, then $\Lambda(\bar{x}) \neq \emptyset$. This shows that any limit point $\bar{x}$ of the sequence generated by the algorithm, that satisfies the constraint qualification, is a KKT point. Not all constraint qualifications may be used for this purpose, but this separated analysis has helped in identifying new weak constraint qualifications suitable for global convergence analysis. See, for instance, [4, 5, 6]. Also, this greatly simplifies the analysis of an algorithm, which resorts to proving some property of the sequence it generates, instead of its limit.

In summary the global convergence of an algorithm using a sequential optimality condition may be done as follows:

a) Characterize the type of sequences $\{x^k\}$ that the algorithm generates with a mathematical proposition $\mathcal{P}(\{x^k\})$.

b) Prove that at a local minimizer $\bar{x}$ of the problem, there exists a sequence $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds.

c) Measure the strength of $\mathcal{P}(\cdot)$ by showing that a point $\bar{x}$, that can be approximated by $z^k \to \bar{x}$ such that $\mathcal{P}(\{z^k\})$ holds, has the property that whenever $\bar{x}$ satisfies some constraint qualification, then a standard first- or second-order necessary optimality condition is satisfied at $\bar{x}$.

In the remainder of this section we proceed with item b), while in the next section we continue with the analysis of items a) and c). This means, we first develop a strong sequential optimality condition and secondly prove that for a local minimizer $\bar{x}$ of (NLP) there exists a sequence $\{z^k\}$ converging to $\bar{x}$ so that $\{z^k\}$ satisfies this optimality condition. Items a) and c) will be dealt with in Section 4 and are related to our main goal of building an algorithm whose limit points satisfy a strong second-order necessary optimality condition, based on the critical cone (2.1), as used in Proposition 2.3, instead of its lineality space (2.4) in Proposition 2.4.

At this point, we do not assume a constraint qualification to hold with respect to the whole feasible set Ω. However, the following constant rank condition with respect to the set of inequality constraints will be used.

Assumption 3.1. It is said that a point $\bar{x} \in \mathbb{R}^n$ satisfies this assumption if there is a neighborhood of $\bar{x}$ so that, for any subset $J \subseteq A(\bar{x})$, the rank of the family $\{\nabla g_j(y)\}_{j \in J}$ is constant for all y in this neighborhood.

Assumption 3.1 can be seen as CRCQ for a feasible point of a constraint set defined by the inequality constraints of (NLP) only. This is clearly not a constraint qualification for (NLP). Notice that Assumption 3.1 holds trivially if the functions g_j are affine. In order to present our definition of a strong second-order sequential optimality condition for problems such that Assumption 3.1 holds, let us consider the perturbed critical cones

$$C_1(y, x, \mu) := \left\{ \quad d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla h_i(y)^\top d = 0 & \text{for } \; i = 1, \ldots, m, \\ \nabla g_j(y)^\top d \leq 0 & \text{for } \; j \in A(x) \text{ with } \; \mu_j = 0, \\ \nabla g_j(y)^\top d = 0 & \text{for } \; j \in A(x) \text{ with } \; \mu_j > 0. \end{array} \right\}$$

and

$$C_2(y, x) := \left\{ \quad d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla f(y)^\top d = 0, & \\ \nabla h_i(y)^\top d = 0 & \text{for } \; i = 1, \ldots, m, \\ \nabla g_j(y)^\top d \leq 0 & \text{for } \; j \in A(x). \end{array} \right\},$$

for $x, y \in \mathbb{R}^n$ and $\mu \in \mathbb{R}_+^p$. Notice that when $(\bar{x}, \lambda, \mu)$ is a KKT point of (NLP), it holds that

$$C_1(\bar{x}, \bar{x}, \mu) = C(\bar{x}) = C_2(\bar{x}, \bar{x})$$

with $C(x)$ defined in (2.1).

Definition 3.1. (Strong-AKKT2) A point $\bar{x}$ satisfies the C_1-Strong Approximate-KKT2 (C_1-SAKKT2) condition for (NLP) if there exists a sequence $(x^k, \lambda^k, \mu^k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+ \times (0, \infty)$ with $x^k \to \bar{x}$ and $\varepsilon_k \searrow 0$ such that

$$\left\| \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k \tag{3.1}$$

$$\|h(x^k)\| \leq \varepsilon_k, \quad \|\max\{0, g(x^k)\}\| \leq \varepsilon_k, \quad \|\min\{\mu^k, -g(x^k)\}\| \leq \varepsilon_k, \tag{3.2}$$

and

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla^2 g_j(x^k) \right) d \geq -\varepsilon_k \|d\|^2 \quad \text{for all } d \in C_1(x^k, \bar{x}, \mu^k). \tag{3.3}$$

If one replaces $C_1(x^k, \bar{x}, \mu^k)$ by $C_2(x^k, \bar{x})$ in the previous definition we say that $\bar{x}$ satisfies the C_2-SAKKT2 condition.

We now prove that C_1-SAKKT2 is a necessary optimality condition for problems which fulfill Assumption 3.1 while C_2-SAKKT2 is a necessary condition for strict optimality under Assumption 3.1.

Theorem 3.2. *Let $\bar{x}$ be a local minimizer of* (NLP) *and suppose that Assumption 3.1 holds at $\bar{x}$. Then $\bar{x}$ satisfies the C_1-SAKKT2 condition. If, in addition, $\bar{x}$ is a strict local minimizer of* (NLP)*, then $\bar{x}$ satisfies the C_2-SAKKT2 condition.*

Proof. Let $\delta > 0$ be chosen such that $f(\bar{x}) \leq f(x)$ holds for all $x \in \Omega$ with $\|x - \bar{x}\| \leq \delta$. Given a sequence $\{\rho_k\} \subset \mathbb{R}_+$ with $\rho_k \to +\infty$, we consider the regularized penalty subproblem, where only the equality constraints are penalized, that is,

$$\begin{aligned} &\text{Minimize} && \phi_k(x) := f(x) + \frac{\rho_k}{2} \sum_{i=1}^{m} h_i(x)^2 + \frac{1}{4}\|x - \bar{x}\|^4, \\ &\text{subject to} && g_j(x) \leq 0, \quad j = 1, \dots, p, \\ &&& \|x - \bar{x}\| \leq \delta. \end{aligned} \tag{3.4}$$

Let x^k be a global solution of the optimization problem (3.4), which exists because its feasible set is nonempty and compact and the objective function is continuous. Therefore, for any $k \in \mathbb{N}$, we have

$$f(x^k) + \frac{1}{4}\|x^k - \bar{x}\|^4 \leq \phi_k(x^k) \leq \phi_k(\bar{x}) = f(\bar{x}). \tag{3.5}$$

Moreover, because $\|x^k - \bar{x}\| \leq \delta$ is valid for all $k \in \mathbb{N}$, there is x^* and an infinite subset $K \subseteq \mathbb{N}$ so that $\lim_{k \in K} x^k = x^*$. Notice that $g(x^*) \leq 0$ and $\|x^* - \bar{x}\| \leq \delta$. Further, since $\rho_k \to +\infty$ and $\{\phi_k(x^k)\}_{k \in K}$ is bounded, we have

$$\lim_{k \in K} h(x^k) = 0 \tag{3.6}$$

so that $h(x^*) = 0$ follows. From (3.5) taken for $k \in K$, we also conclude that

$$f(x^*) + \frac{1}{4}\|x^* - \bar{x}\|^4 \leq f(\bar{x}).$$

This, $g(x^*) \leq 0$, $h(x^*) = 0$, $\|x^* - \bar{x}\| \leq \delta$, and the definition of δ imply $f(\bar{x}) \leq f(x^*)$ so that $x^* = \bar{x}$ follows. Then, for $k \in K$ large enough, we have that $\|x^k - \bar{x}\| < \delta$, i.e., the constraint $\|x - \bar{x}\| \leq \delta$ in (3.4) is not active at $x = x^k$ for these $k \in K$. Hence, applying Proposition 2.3 (with (NLP) replaced by problem (3.4)) for each of these large enough $k \in K$, it follows by Assumption 3.1 that there exists a Lagrange multiplier $\mu^k \in \mathbb{R}^p_+$ such that (x^k, μ^k) is a KKT point of (3.4) that satisfies a strong second-order necessary optimality condition. More in detail, we have that

$$g(x^k) \leq 0, \quad \mu^k \geq 0, \quad g(x^k)^\top \mu^k = 0, \tag{3.7}$$

$$\begin{aligned} \nabla_x L_{(3.4)}(x^k, \mu^k) &:= \nabla \phi_k(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla h_i(x^k) + \|x^k - \bar{x}\|^2 (x^k - \bar{x}) \\ &\quad + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \\ &= 0, \end{aligned} \tag{3.8}$$

and, because of

$$\begin{aligned}\nabla^2_{xx}L_{(3.4)}(x^k,\mu^k) &= \nabla^2 f(x^k)+\rho_k\sum_{i=1}^m h_i(x^k)\nabla^2 h_i(x^k)+\nabla h_i(x^k)\nabla h_i(x^k)^\top\\ &\quad+\sum_{j=1}^p \mu_j^k\nabla^2 g_j(x^k)+2(x^k-\bar{x})(x^k-\bar{x})^\top+\|x^k-\bar{x}\|^2 I,\end{aligned}$$

we further have that

$$d^\top\nabla^2_{xx}L_{(3.4)}(x^k,\mu^k)d\geq 0 \quad \text{for all } d\in C_{(3.4)}(x^k) \tag{3.9}$$

with

$$C_{(3.4)}(x^k):=\left\{d\in\mathbb{R}^n \mid \nabla\phi_k(x^k)^\top d\leq 0, \nabla g_j(x^k)^\top d\leq 0 \text{ for all } j\in A(x^k)\right\}.$$

Since (x^k,μ^k) satisfies (3.7) and (3.8), this yields

$$\begin{aligned}C_{(3.4)}(x^k) &= \left\{d\in\mathbb{R}^n \,\middle|\, \sum_{j=1}^p \mu_j^k\nabla g_j(x^k)^\top d=0, \nabla g_j(x^k)^\top d\leq 0 \text{ for } j\in A(x^k)\right\}\\ &= \left\{d\in\mathbb{R}^n \,\middle|\, \begin{array}{l}\nabla g_j(x^k)^\top d\leq 0 \text{ for } j\in A(x^k) \text{ with } \mu_j^k=0,\\ \nabla g_j(x^k)^\top d=0 \text{ for } j\in A(x^k) \text{ with } \mu_j^k>0\end{array}\right\}.\end{aligned}$$

Defining

$$\lambda^k:=\rho_k h(x^k), \quad \varepsilon_k:=\max\{\|x^k-\bar{x}\|,\|h(x^k)\|\} \quad \text{for } k\in K,$$

we first obtain that $\varepsilon_k\searrow 0$ for $k\in K$. Without loss of generality, we assume that $k\in K$ is large enough so that, due to (3.7), $\mu_j^k=0$ for $j\notin A(\bar{x})$. Thus, it follows from (3.8) and (3.7) that, for $k\in K$ sufficiently large,

$$\left\|\nabla f(x^k)+\sum_{i=1}^m \lambda_i^k\nabla h_i(x^k)+\sum_{j=1}^p \mu_j^k\nabla g_j(x^k)\right\|\leq\|x^k-\bar{x}\|^3\leq\varepsilon_k,$$

$$\|h(x^k)\|\leq\varepsilon_k, \quad \|\max\{0,g(x^k)\}\|=0$$

by $g_j(x^k)\leq 0$ according to (3.4), and

$$\min\{\mu^k,-g(x^k)\}=0.$$

Therefore, the requirements (3.1) and (3.2) in Definition 3.1 are satisfied. Furthermore, since $A(x^k)\subseteq A(\bar{x})$ for $k\in K$ sufficiently large, we have

$$C_1(x^k,\bar{x},\mu^k)\subseteq C_{(3.4)}(x^k)\cap\{d\in\mathbb{R}^n \mid \nabla h(x^k)^\top d=0\}.$$

Taking any $d\in C_1(x^k,\bar{x},\mu^k)$, we further get from (3.9) with the definitions of λ^k and ε_k that, for $k\in K$ sufficiently large,

$$\begin{aligned}d^\top\nabla^2_{xx}L(x^k,\lambda^k,\mu^k)d &= d^\top\left(\nabla^2 f(x^k)+\sum_{i=1}^m\lambda_i^k\nabla^2 h_i(x^k)+\sum_{i\in A(\bar{x})}\mu_i^k\nabla^2 g_i(x^k)\right)d\\ &\geq -d^\top\left(2(x^k-\bar{x})(x^k-\bar{x})^\top-\|x^k-\bar{x}\|^2 I\right)d\\ &\geq -\varepsilon_k\|d\|^2.\end{aligned}$$

Hence, also (3.3) in Definition 3.1 holds and, altogether, $\bar{x}$ satisfies the C_1-SAKKT2 condition.

Assume now that $\bar{x}$ is a strict local minimizer of (NLP). Thus, we can follow exactly the same proof with $\phi_k(x)$ replaced by $\tilde{\phi}_k(x):=f(x)+\frac{\rho_k}{2}\sum_{i=1}^m h_i(x)^2$. Note that the expression for $C_{(3.4)}(x^k)\cap\{d\in\mathbb{R}^n \mid \nabla h(x^k)^\top d=0\}$, again with ϕ_k replaced by $\tilde{\phi}_k$, contains $C_2(x^k,\bar{x})$, which concludes the proof. □

4 Generating KKT Points with Strong Second-Order Conditions

Typically, second-order algorithms are only shown to generate a sequence that converges to a point that satisfies the weak second-order necessary optimality condition from Proposition 2.4. In [12], it is shown that limit points of a standard barrier method need not satisfy the stronger second-order necessary optimality condition from Proposition 2.3, even if strict local minimizers for the subproblems are found at each iteration. In detail, the authors from [12] considered the counterexample

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}x^\top Hx, \\ \text{subject to} & x \geq 0, \end{array}$$

where $x \in \mathbb{R}^n$ with $n \geq 4$ and $H = I - \frac{3}{2n(n-1)}zz^\top$ with $z = e - ne_1$, where e_1 is the first canonical vector and e is the vector with 1 in all entries. For any sequence $r_k \searrow 0$, let $x^k = \sqrt{r_k}e \to \bar{x} = 0$ be defined. Notice that $\mu = 0$ is the unique Lagrange multiplier associated with $\bar{x}$, that is, $\nabla_x L(\bar{x}, \mu) = 0$. However, one has

$$e_1^\top \nabla^2_{xx} L(\bar{x}, \mu)e_1 = e_1^\top He_1 = 1 - \frac{3(n-1)}{2n} < 0 \quad \text{with } e_1 \in C(\bar{x}) = \mathbb{R}^n_+.$$

Thus, the sequence $\{x^k\}$ converges to a point that fails to satisfy the strong second-order necessary optimality condition. Note however that x^k is a strict local minimizer of the barrier function subproblem

$$\begin{array}{ll} \text{Minimize} & b(x, r_k) := \frac{1}{2}x^\top Hx - r_k \sum_{i=1}^n \log(x_i) \\ \text{subject to} & x > 0. \end{array}$$

Indeed, one has

$$\nabla_x b(x^k, r_k) = Hx^k - r_k(x^k)^{-1} = 0 \quad \text{and} \quad \nabla^2_{xx} b(x^k, r_k) = H + r_k(x^k)^{-2} = \frac{1}{2}I + \frac{3}{2}\left(I - \frac{zz^\top}{z^\top z}\right),$$

where the latter is clearly positive definite. Here, $(x^k)^{-1}$ and $(x^k)^{-2}$ were used to denote, respectively, the componentwise inverse vector and the diagonal matrix with inverse-squared diagonal entries of x^k as defined above. The same example from [12] was analyzed in [7]. There, it was shown that a second-order augmented Lagrangian method may also generate the same sequence x^k as above in such a way that x^k is a strict local minimizer of the corresponding augmented Lagrangian subproblems

$$\text{Minimize} \quad \frac{1}{2}x^\top Hx + \rho_k \sum_{i=1}^n \max\left\{0, -x_i + \frac{\mu_i^k}{\rho_k}\right\}^2$$

for standard approximate Lagrange multipliers μ^k and penalty parameters ρ_k.

These results suggest that in order to generate points satisfying a stronger second-order necessary condition for (NLP), one should not penalize inequality constraints. Therefore, we consider the simple penalty algorithm below whose subproblems penalize only equality constraints, while the inequality constraints are kept within the subproblems.

To define the subproblems later on, let $\rho > 0$ be given and consider the problem

$$\begin{array}{ll} \text{Minimize} & F_\rho(x) := f(x) + \frac{1}{2}\rho\|h(x)\|^2, \\ \text{subject to} & g(x) \leq 0. \end{array} \tag{4.1}$$

Proposition 4.1. *Suppose that a local minimizer x_ρ of (4.1) is strict and satisfies Assumption 3.1. Then, for any $\varepsilon > 0$, there exist $x = x(\rho, \varepsilon) \in \mathbb{R}^n$ and $\mu = \mu(\rho, \varepsilon) \in \mathbb{R}^p_+$ which solve the $KKT(\rho, \varepsilon)$ system given by*

$$\left\| \nabla F_\rho(x) + \sum_{j=1}^p \mu_j \nabla g_j(x) \right\| \leq \varepsilon, \tag{4.2}$$

$$\|\max\{0, g(x)\}\| \leq \varepsilon, \quad \|\min\{\mu, -g(x)\}\| \leq \varepsilon,$$

$$d^\top \left(\nabla^2 F_\rho(x) + \sum_{j=1}^p \mu_j \nabla^2 g_j(x) \right) d \geq -\varepsilon\|d\|^2 \quad \text{for all } d \in C_\rho(x),$$

where

$$C_\rho(x) := \left\{ d \in \mathbb{R}^n \mid \nabla F_\rho(x)^\top d = 0,\ \nabla g_j(x)^\top d \leq 0 \text{ for all } j \in A(x) \right\}.$$

Proof. The proposition follows immediately by applying Theorem 3.2 to problem (4.1) because Assumption 3.1 is requested to hold at the strict local minimizer x_ρ of problem (4.1). ☐

Based on the KKT(ρ, ε) system above, we consider the following simple algorithm.

Algorithm 1

Let sequences $\{\varepsilon_k\}$, $\{\rho_k\} \subset (0, \infty)$ with $\varepsilon_k \searrow 0$ and $\rho_k \to \infty$ be given. Set $k := 0$.

Step 1: Compute $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^p_+$ as a solution of $KKT(\rho_k, \varepsilon_k)$.

Step 2: Set $k := k + 1$ and go back to Step 1.

Our global convergence result will partly rely on the following assumption, which is related to Assumption 3.1.

Assumption 4.1. It is said that a point $\bar{x} \in \mathbb{R}^n$ satisfies this assumption if there is a neighborhood of $\bar{x}$ so that, for any subset $J \subseteq A(\bar{x})$, the rank of the family $\{\nabla f(y)\} \cup \{\nabla h_i(y)\}_{i=1}^m \cup \{\nabla g_j(y)\}_{j\in J}$ is constant for all y in this neighborhood.

Moreover, in part b) of Theorem (4.2), we will employ the extended MFCQ, which is satisfied at a point $\bar{x} \in \mathbb{R}^n$, if

- the column rank of $\nabla h(\bar{x}) \in \mathbb{R}^{n\times m}$ is equal to m and
- there is $d \in \mathbb{R}^n$ such that $\nabla h(\bar{x})^\top d = 0$ and $\nabla g_j(\bar{x})^\top d < 0$ for all $j \in A(\bar{x})$.

Note that the extended MFCQ does not require that $\bar{x}$ belongs to the feasible set Ω. However, in the theorem below, $g(\bar{x}) \leq 0$ holds by construction.

Theorem 4.2. *Let $\{(x^k, \mu^k)\}$ be an infinite sequence generated by Algorithm 1. Further assume that the sequence $\{x^k\}$ has an accumulation point $\bar{x}$. Then, the following assertions hold:*

a) If $h(\bar{x}) = 0$, then $\bar{x}$ satisfies the C_2-SAKKT2 condition.

b) If the extended MFCQ and Assumption 4.1 are satisfied at $\bar{x}$, then there are $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ so that $(\bar{x}, \lambda, \mu)$ fulfills the KKT conditions of (NLP) and the second-order condition $\langle d, \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d\rangle \geq 0$ for all $d \in C(\bar{x})$.

Proof. a) Let us assume without loss of generality that $x^k \to \bar{x}$. Note that

$$\begin{aligned} \nabla F_{\rho_k}(x^k) &= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k), \\ \nabla^2 F_{\rho_k}(x^k) &= \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \rho_k \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i(x^k)^\top \end{aligned} \tag{4.3}$$

with $\lambda_i^k := \rho_k h_i(x^k)$ for $i = 1, \ldots, m$. Thus, by Step 1 in Algorithm 1, it follows that

$$\left\| \nabla F_{\rho_k}(x^k) + \sum_{i=1}^p \mu_j^k \nabla g_j(x^k) \right\| = \left\| \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^p \mu_j^k \nabla g_j(x^k) \right\| \leq \varepsilon_k, \tag{4.4}$$

$$\| \max\{0, g(x^k)\}\| \leq \varepsilon_k, \quad \| \min\{\mu^k, -g(x^k)\}\| \leq \varepsilon_k, \tag{4.5}$$

and, having (4.3) in mind,

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) \right) d \geq -\varepsilon_k \|d\|^2 - \rho_k d^\top \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i(x^k)^\top d, \tag{4.6}$$

for all $d \in C_{\rho_k}(x^k)$. Since $h(\bar{x}) = 0$ is assumed in assertion a), we have

$$\lim_{k\to\infty} \|h(x^k)\| = 0. \tag{4.7}$$

To complete the proof that $\bar{x}$ satisfies the C_2-SAKKT2 condition, we observe that $A(x^k) \subseteq A(\bar{x})$ for k sufficiently large and, as a consequence,

$$\begin{aligned} C_2(x^k, \bar{x}) &= \{ d \in \mathbb{R}^n \mid \nabla f(x^k)^\top d = 0, \nabla h_i(x^k)^\top d = 0 \text{ for } i = 1, \ldots, m, \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(\bar{x}) \} \\ &\subseteq \left\{ d \in \mathbb{R}^n \mid \nabla f(x^k)^\top d + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k)^\top d = 0, \nabla g_j(x^k)^\top d \leq 0 \text{ for } j \in A(x^k) \right\} \\ &= C_{\rho_k}(x^k) \end{aligned}$$

is valid for k sufficiently large. According to this and (4.3), (4.6) yields

$$d^\top \left(\nabla^2 f(x^k) + \sum_{i=1}^m \lambda_j^k \nabla^2 h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla^2 g_j(x^k) \right) d \geq -\varepsilon_k \|d\|^2 \quad \text{for all } d \in C_2(x^k, \bar{x})$$

for all sufficiently large k. This, (4.4), (4.5), and (4.7) show that $\bar{x}$ satisfies the C_2-SAKKT2 condition.

b) Since the extended MFCQ is assumed to hold at $\bar{x}$, it is well known that $\{(\lambda^k, \mu^k)\}$ is bounded. Indeed, if this would be not the case, we can divide formula (4.4) by $\|(\lambda^k, \mu^k)\|$. Then an infinite index set $K \subseteq \mathbb{N}$ exists with

$$\lim_{k\in K} \frac{(\lambda^k, \mu^k)}{\|(\lambda^k, \mu^k)\|} = (\alpha, \beta) \neq 0 \quad \text{and } \beta \geq 0.$$

Taking the limit in (4.4) and (4.5), we obtain

$$\sum_{i=1}^{m} \alpha_i \nabla h_i(\bar{x}) + \sum_{j=1}^{p} \beta_j \nabla g_j(\bar{x}) = 0, \quad g(\bar{x}) \le 0, \quad \text{and} \quad \min\{\beta, -g(\bar{x})\} = 0, \tag{4.8}$$

which leads to $(\alpha, \beta) = 0$ due to the extended MFCQ. This contradicts $\|(\alpha, \beta)\| = 1$. Hence, for an infinite index set $K_1 \subseteq K$, we have that

$$\lim_{k \in K_1} (\lambda^k, \mu^k) = (\lambda, \mu) \in \Lambda(\bar{x}).$$

Therefore, $\lambda_i^k = \rho_k h_i(x^k)$ for $i = 1, \ldots, m$ and $\rho_k \to \infty$ imply

$$\lim_{k \in K_1} \|h(x^k)\| = 0,$$

i.e., by the continuity of h, it follows that $h(\bar{x}) = 0$. Moreover, according to (4.8), we also have $g(\bar{x}) \le 0$. Thus, $\bar{x} \in \Omega$ so that $(\bar{x}, \lambda, \mu)$ is a KKT point of (NLP).

To complete the proof of part b), take any

$$d \in C(\bar{x}) = \{d \in \mathbb{R}^n \mid \nabla f(\bar{x})^\top d \le 0,\ \nabla h(\bar{x})^\top d = 0, \nabla g_j(\bar{x})^\top d \le 0 \text{ for all } j \in A(\bar{x})\}$$

with $C(x)$ defined in (2.1). Since $(\bar{x}, \lambda, \mu)$ is a KKT point, we easily see that $\nabla f(\bar{x})^\top d = 0$. Now, let $J \subseteq A(\bar{x})$ denote the set of all indexes such that $\nabla g_j(\bar{x})^\top d = 0$. By Assumption 4.1 and using the proof technique in [1, Lemma 3.1], we get that there exists a sequence $d^k \to d$ such that

$$\nabla f(x^k)^\top d^k = 0, \quad \nabla h_i(x^k)^\top d^k = 0 \text{ for } i = 1, \ldots, m, \quad \text{and} \quad \nabla g_j(x^k)^\top d^k = 0 \text{ for } j \in J.$$

Since $\nabla g_j(\bar{x})^\top d < 0$ for $j \notin J$, we have for k large enough that $d^k \in C_2(x^k, \bar{x})$. Using direction d^k in (3.3) with $C_1(x^k, \bar{x}, \mu^k)$ replaced by $C_2(x^k, \bar{x})$, we may take the limit for $k \to \infty$ and get $\langle d, \nabla^2_{xx} L(\bar{x}, \lambda, \mu) d\rangle \ge 0$. As this can be done for all $d \in C(\bar{x})$, the proof is complete. □

We end by noting that Assumption 4.1 cannot be removed in the previous result. Indeed, let us consider a modification of the example given by Baccari in [8] and let us apply Algorithm 1.

Example 4.3. For the problem

$$\begin{array}{ll} \text{Minimize} & x_3, \\ \text{subject to} & x_3 \ge 2\sqrt{3}x_1x_2, \\ & x_3 \ge x_2^2 - 3x_1^2, \\ & x_3 \ge -2\sqrt{3}x_1x_2 - 2x_2^2, \\ & x_3 = 0, \end{array}$$

the point $\bar{x} = (0, 0, 0)$ is a global minimizer that satisfies MFCQ. Take a sequence $\rho_k \to \infty$ and consider the sequence of subproblems as associated to $\{\rho_k\}$ by Algorithm 1, i.e., just the equality constraint $x_3 = 0$ is penalized and the inequality constraints are kept within the subproblems.Thus, the subproblems read as follows:

$$\begin{array}{ll} \text{Minimize} & x_3 + \frac{\rho_k}{2}x_3^2, \\ \text{subject to} & x_3 \ge 2\sqrt{3}x_1x_2, \\ & x_3 \ge x_2^2 - 3x_1^2, \\ & x_3 \ge -2\sqrt{3}x_1x_2 - 2x_2^2. \end{array} \tag{4.9}$$

We take the constant sequence $x^k = \bar{x}$ for all k, since $\bar{x}$ is the global solution for every k. However, it can easily be calculated that the point $\bar{x}$ does not satisfy the strong second-order necessary optimality condition (2.3).

References

[1] R. Andreani, J.M. Martínez and M.L. Schuverdt, On second-order optimality conditions for nonlinear programming, *Optimization* 56 (2007) 529–542.

[2] R. Andreani, C.E. Echagüe and M.L. Schuverdt, Constant-Rank Condition and Second-Order Constraint Qualification, *J. Optim. Theory Appl.* 146 (2010) 255–266.

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

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

[5] R. Andreani, G. Haeser, M.L. Schuverdt and P.J.S. Silva, Two new weak constraint qualifications and applications, *SIAM J. Optim.* 22 (2012) 1109–1135.

[6] R. Andreani, J.M. Martínez, A. Ramos and P.J.S. Silva, A cone-continuity constraint qualification and algorithmic consequences, *SIAM J. Optim.* 26 (2016) 96–110.

[7] R. Andreani and L.D. Secchin, A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points, *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 6:1 (2018) 1–7.

[8] A. Baccari, On the classical necessary second-order optimality conditions, *J. Optim. Theory Appl.* 123 (2004) 213–221.

[9] R. Behling, G. Haeser, A. Ramos and D.S. Viana, On a conjecture in second-order optimality conditions, *J. Optim. Theory Appl.* 176 (2018) 625–633. Extended version at arXiv:1706.07833.

[10] A. Ben-Tal, Second-order and related extremality conditions in nonlinear programming, *J. Optim. Theory Appl.* 31 (1980) 143–165.

[11] H. Gfrerer and B.S. Mordukhovich, Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions, *SIAM J. Optim.* 25 (2015) 2081–2119.

[12] N.I.M. Gould and P.L. Toint, A note on the convergence of barrier algorithms to second-order necessary points, *Math. Program.* 85 (1999) 433–438.

[13] G. Haeser, An extension of Yuan's Lemma and its applications in optimization, *J. Optim. Theory Appl.* 174 (2017) 641–649.

[14] G. Haeser, H. Liu and Y. Ye, Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary, *Math. Program.* 178 (2019) 263–299.

[15] R. Janin, Direction derivate of the marginal function in nonlinear programming, *Mathematical Programming Study* 21 (1984) 110–126.

[16] W. Mascarenhas, A simple canonical form for nonlinear programming problems and its use, *J. Optim. Theory Appl.* 181 (2019) 456–469.

[17] K.G. Murty and S.N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming, *Math. Program.* 39 (1987) 117–129.

[18] M. O'Neill and S.J. Wright, A log-barrier Newton-CG method for bound constrained optimization with complexity guarantees, *IMA J. Numer. Anal.* 41 (2021) 84–121.

Andreas Fischer
Faculty of Mathematics, Institute of Numerical Mathematics
Technische Universität Dresden, 01062 Dresden, Germany
E-mail address: andreas.fischer@tu-dresden.de

Gabriel Haeser
Department of Applied Mathematics
University of São Paulo, São Paulo-SP, Brazil
E-mail: ghaeser@ime.usp.br

Thiago P. Silveira
Department of Applied Mathematics
University of São Paulo, São Paulo-SP, Brazil
E-mail: thiagops@ime.usp.br