ON THE ACCURATE IDENTIFICATION OF ACTIVE CONSTRAINTS *

FRANCISCO FACCHINEI†, ANDREAS FISCHER‡, AND CHRISTIAN KANZOW§

Abstract. We consider nonlinear programs with inequality constraints, and we focus on the problem of identifying those constraints which will be active at an isolated local solution. The correct identification of active constraints is important from both a theoretical and a practical point of view. Such an identification removes the combinatorial aspect of the problem and locally reduces the inequality constrained minimization problem to an equality constrained one which can be more easily dealt with. We present a new technique which identifies active constraints in a neighborhood of a solution and which requires neither complementary slackness nor uniqueness of the multipliers. We also present extensions to variational inequalities and numerical examples illustrating the identification technique.

Key words. Constrained optimization, variational inequalities, active constraints, degeneracy, identification of active constraints.

AMS subject classifications. 90C30, 65K05, 90C33, 90C31.

1. Introduction. In this paper we consider the problem of identifying the constraints which are active at an isolated stationary point of the nonlinear program

$$\min \quad f(x) \quad \text{s.t.} \quad g(x) \geq 0, \qquad (\mathcal{P})$$

where it is assumed that the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are at least continuously differentiable. More specifically, we are interested in the following question: Given an $(x, \lambda) \in \mathbb{R}^{n+m}$ belonging to a sufficiently small neighborhood of a Karush-Kuhn-Tucker (KKT) point $(\bar{x}, \bar{\lambda})$ of Problem $(\mathcal{P})$, is it possible to correctly estimate, on the basis of the problem data in x, the set of indices

$$I_0 := \{i |\, g_i(\bar{x}) = 0\}$$

of the active constraints? The correct identification of active constraints is important from both a theoretical and a practical point of view. Such an identification, by removing the difficult combinatorial aspect of the problem, locally reduces the inequality constrained minimization problem to an equality constrained one which is much easier to deal with. In particular, the study of the local convergence rate of most algorithms for Problem $(\mathcal{P})$ implicitly or explicitly depends on the fact that I_0 is eventually identified.

Theoretically, the identification of the active constraints is not difficult if strict complementarity holds at the solution, see the discussion in the next section. However, as far as we are aware of, to date no technique can successfully identify all active constraints if the strict complementary slackness assumption is violated, except in the case of linear (complementarity) problems, see [10, 11, 21]. In this paper we present a new technique which, under mild assumptions, correctly identifies active

*The work of the authors was partially supported by NATO under grant CRG 960137.

†Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Buonarroti 12, I-00185 Roma, Italy (soler@dis.uniroma1.it).

‡Department of Mathematics, University of Dortmund, D-44221 Dortmund, Germany (fischer@math.uni-dortmund.de).

§Institute of Applied Mathematics, University of Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany (kanzow@math.uni-hamburg.de).

constraints in a neighborhood of a KKT point. This technique appears to improve on existing techniques. In particular, it enjoys the following properties:

(i) It is simple and independent of the algorithm used to generate the point (x, λ).
(ii) It does not require strict complementary slackness.
(iii) It does not require uniqueness of the multipliers.
(iv) It is able to handle problems with nonlinear constraints.
(v) It does not rely on any convexity assumption.
(vi) In the case of unique multipliers it also permits the correct identification of strongly active constraints.
(vii) The identification technique can be applied also to the Karush-Kuhn-Tucker system arising from variational inequalities.

Strategies for identifying active constraints are part of the optimization folklore [2, 15, 17], however, they almost invariably lack some of the good characteristics listed above. In the last ten years a special attention has been devoted to this problem in the field of interior point methods for linear (complementarity) problems [10, 11, 21], where satisfactory results have been reached. Recent works on the nonlinear case include [9, 12, 28], where the case of box constraints is considered, and [13, 41, 42], where the general nonlinear case is studied. Related material can also be found in [4, 5, 6], where the problem of establishing whether a sequence $\{x^k\}$ converging to a solution $\bar{x}$, in some way, eventually identifies the set I_0 is dealt with.

We remark that, in order to identify the active set, we suppose we are given a pair (x, λ) of primal and dual variables. If we think of algorithmic applications of the results in this paper, we stress that most algorithms will produce a sequence of primal and dual variables. Even in the rare cases in which this does not occur, it is usually possible, under reasonable assumptions, to generate a continuous dual estimate by using a *multiplier function*, see, e.g., [13] and references therein.

This paper is organized as follows. In the next section we introduce the identification technique and prove its main properties. The identification technique critically depends on the definition of what we call an *identification function*. Therefore, the more technical Section 3 is devoted to the definition of identification functions under different sets of assumptions. In Section 4 we give some numerical examples and in Section 5 we make some final comments.

We conclude this section by providing a list of the notation employed. Throughout the paper, $\|\cdot\|$ indicates the Euclidean vector norm. The symbol B_ϵ denotes the open Euclidean ball with radius $\epsilon > 0$ and center at the origin; the dimension of the space will be clear from the context. The Euclidean distance of a point y from a nonempty set S is abbreviated by $\mathrm{dist}[y, S]$. We write x_+ for the vector $\max\{0, x\}$, where the maximum is taken componentwise. We set $I := \{1, \dots, m\}$ and make use of the notation x_J for $J \subseteq I$ in order to represent the $|J|$-dimensional vector with components $x_i, i \in J$. Finally, the transposed Jacobian of the vector-valued mapping g at a point x will be denoted by $\nabla g(x)$, i.e., the ith column of this matrix is the gradient $\nabla g_i(x)$.

2. Identifying Active Constraints. Following the usual terminology in constrained optimization, we call a vector $\bar{x} \in \mathbb{R}^n$ a *stationary point* of $(\mathcal{P})$ if there exists a vector $\bar\lambda \in \mathbb{R}^m$ such that $(\bar x, \bar\lambda)$ solves the *Karush-Kuhn-Tucker system*

$$\begin{array}{rcl} \nabla f(x) - \nabla g(x)\lambda & = & 0, \\ \lambda & \geq & 0, \\ g(x) & \geq & 0, \\ \lambda^T g(x) & = & 0. \end{array} \tag{2.1}$$

The pair $(\bar x, \bar\lambda)$ is called a *KKT point* of Problem $(\mathcal{P})$. In the sequel $\bar x$ will always denote a fixed, isolated stationary point, so that there is a neighborhood of $\bar x$ which does not contain any further stationary point of $(\mathcal{P})$. Moreover, we shall indicate by Λ the set of all Lagrange multipliers $\bar\lambda$ associated with $\bar x$ and by $\mathcal{K}$ the set of all KKT points associated with $\bar x$, that is,

$$\Lambda := \{\bar\lambda \,|\, (\bar x, \bar\lambda) \text{ solves (2.1)}\}, \qquad \mathcal{K} := \{(\bar x, \bar\lambda) \,|\, \bar\lambda \in \Lambda\}.$$

The set Λ is closed and convex and therefore, so is the set $\mathcal{K}$. Gauvin [16] showed that Λ is bounded (and hence compact) if and only if the *Mangasarian-Fromovitz constraint qualification* (MFCQ) is satisfied, i.e., if and only if

$$\sum_{i\in I_0} u_i \nabla g_i(\bar x) = 0, \quad u_i \geq 0 \ \forall i \in I_0 \qquad \Longrightarrow \qquad u_i = 0 \ \forall i \in I_0.$$

On the other hand, Kyparisis [27] showed that Λ reduces to a singleton if and only if the *strict Mangasarian-Fromovitz constraint qualification* (SMFCQ) holds, i.e., if and only if

$$\sum_{i\in I_0} u_i \nabla g_i(\bar x) = 0, \quad u_i \geq 0 \ \forall i \in I_0 \setminus I_+ \qquad \Longrightarrow \qquad u_i = 0 \ \forall i \in I_0,$$

where I_+ denotes the index set

$$I_+ := \{i \in I_0 \,|\, \exists \bar\lambda \in \Lambda : \bar\lambda_i > 0\}.$$

In particular, the *linear independence constraint qualification* (LICQ), i.e., the linear independence of the gradients of the active constraints, implies that Λ is a singleton.

Our basic aim is to construct a rule which is able to assign to every point (x, λ) an estimate $A(x, \lambda) \subseteq I$ so that $A(x, \lambda) = I_0$ holds if (x, λ) lies in a suitably small neighborhood of a point $(\bar x, \bar\lambda) \in \mathcal{K}$.

Usually estimates of this kind are obtained by comparing the value of $g_i(x)$ to the value of λ_i. For example, it can easily be shown that the set

$$I_\oplus(x, \lambda) := \{i \in I \,|\, g_i(x) \leq \lambda_i\}$$

coincides with the set I_0 for all (x, λ) in a sufficiently small neighborhood of a KKT point $(\bar x, \bar\lambda)$ which satisfies the strict complementarity condition. If this condition is violated, then only the inclusion

$$I_\oplus(x, \lambda) \subseteq I_0 \tag{2.2}$$

holds. Furthermore, if Λ is a singleton, then we also have, in a sufficiently small neighborhood of $(\bar x, \bar\lambda)$,

$$I_+ \subseteq I_\oplus(x, \lambda) \subseteq I_0. \tag{2.3}$$

This relation was exploited to construct locally superlinearly convergent QP-free optimization algorithms when the unique multiplier $\bar{\lambda}$ does not satisfy the strict complementarity condition, see, e.g., [13, 25, 40].

We refer the reader to [2, 13] and references therein for a more complete discussion of these kind of results. An analysis of results established in the literature shows that this conclusion holds in general: if strict complementarity is satisfied, it is usually possible to correctly identify the active constraint set, otherwise a relation like (2.3) is the best result that has been established in the general nonlinear case.

To overcome this situation we propose to compare $g_i(x)$ to a quantity which goes to 0 at a known rate if (x, λ) converges to a point in the KKT set $\mathcal{K}$. To this end, we introduce the notion of an *identification function.*

DEFINITION 2.1. A function $\rho : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ is called an *identification function* for $\mathcal{K}$ if
(a) ρ is continuous on $\mathcal{K}$,
(b) $(\bar{x}, \bar{\lambda}) \in \mathcal{K}$ implies $\rho(\bar{x}, \bar{\lambda}) = 0$,
(c) if $(\bar{x}, \bar{\lambda})$ belongs to $\mathcal{K}$, then

$$\lim_{\substack{(x,\lambda)\to(\bar{x},\bar{\lambda})\\ (x,\lambda)\notin\mathcal{K}}} \frac{\rho(x,\lambda)}{\operatorname{dist}[(x,\lambda),\mathcal{K}]} = +\infty. \tag{2.4}$$

In the next section we shall give examples of how to build, under appropriate assumptions, identification functions. Basically, Definition 2.1 says that a function is an identification function if it goes to 0 when approaching the set $\mathcal{K}$ at a "slower" rate than the distance from the set $\mathcal{K}$. We note that $\operatorname{dist}[(x, \lambda), \mathcal{K}] > 0$ whenever $(x, \lambda) \notin \mathcal{K}$ since $\mathcal{K}$ is a closed set; hence the denominator in (2.4) is always nonzero.

Using Definition 2.1 it is easy to prove that the index set

$$A(x, \lambda) := \{i \in I \mid g_i(x) \le \rho(x, \lambda)\} \tag{2.5}$$

correctly identifies all active constraints if (x, λ) is sufficiently close to the KKT set $\mathcal{K}$.

THEOREM 2.2. *Let ρ be an identification function for $\mathcal{K}$. Then, for any $\bar{\lambda} \in \Lambda$, an $\epsilon = \epsilon(\bar{\lambda}) > 0$ exists such that*

$$A(x, \lambda) = I_0 \qquad \forall (x, \lambda) \in \{(\bar{x}, \bar{\lambda})\} + B_\epsilon. \tag{2.6}$$

Proof. Since g is continuously differentiable, g is locally Lipschitz-continuous. Hence there exists a constant $c > 0$ such that, for all x sufficiently close to $\bar{x}$,

$$g_i(x) \le g_i(\bar{x}) + c\|x - \bar{x}\| \qquad (i \in I). \tag{2.7}$$

Suppose now that $g_i(\bar{x}) = 0$. Then, using (2.4) and (2.7), we obviously have, for $(x, \lambda) \notin \mathcal{K}$ in a sufficiently small neighborhood of $(\bar{x}, \bar{\lambda})$,

$$g_i(x) \le c\|x - \bar{x}\| \le c \operatorname{dist}[(x, \lambda), \mathcal{K}] \le \rho(x, \lambda),$$

so that, by (2.5), $i \in A(x, \lambda)$.

If, instead, $(x, \lambda) \in \mathcal{K}$, then we have $x = \bar{x}$ by the local uniqueness of $\bar{x}$. From the definition of an identification function, we also have $\rho(x, \lambda) = 0$, so that

$$g_i(x) = g_i(\bar{x}) = 0 \le \rho(x, \lambda),$$

and also in this case $i \in A(x, \lambda)$.

On the other hand, if $g_i(\bar{x}) > 0$, it follows, by continuity, that $i \notin A(x, \lambda)$ if (x, λ) is sufficiently close to $(\bar{x}, \bar{\lambda})$. Therefore, for any $\bar{\lambda} \in \Lambda$, we can find $\epsilon = \epsilon(\bar{\lambda}) > 0$ such that (2.6) is satisfied. □

From the previous theorem it is obvious that there exists an open set containing $\mathcal{K}$ where the identification of the active constraints is correct. Using the MFCQ condition we can obtain a somewhat stronger result.

THEOREM 2.3. *Let ρ be an identification function for $\mathcal{K}$. If the MFCQ condition holds, then there is an $\epsilon > 0$ such that*

$$A(x, \lambda) = I_0 \qquad \forall (x, \lambda) \in \mathcal{K} + B_\epsilon.$$

Proof. By the previous Theorem, for every $(\bar{x}, \bar{\lambda}) \in \mathcal{K}$, there exists a neighborhood $\Omega(\epsilon(\bar{\lambda})) = \{(\bar{x}, \bar{\lambda})\} + B_{\epsilon(\bar{\lambda})}$ such that $A(x, \lambda) = I_0$ for every $(x, \lambda) \in \Omega(\epsilon(\bar{\lambda}))$. The collection of open sets $\Omega(\epsilon(\bar{\lambda}))$, $\bar{\lambda} \in \Lambda$, obviously forms an open cover of $\mathcal{K}$. Since the set $\mathcal{K}$ is compact in view of the MFCQ condition, we can extract from the infinite cover $\Omega(\epsilon(\bar{\lambda}))$ with $\bar{\lambda}$ such that $(\bar{x}, \bar{\lambda}) \in \mathcal{K}$ a finite subcover $\Omega(\epsilon(\bar{\lambda}_j))$, with $j = 1, \ldots, s$. Then it is easy to see that the Theorem holds with $\epsilon := \min_{j=1,\ldots,s}\{\epsilon(\bar{\lambda}_j)\}$. □

If the SMFCQ holds, it is even possible to identify the set of *strongly active constraints* at $\bar{x}$, i.e., the set of constraints whose multipliers are positive. To this end, let $A_+(x, \lambda)$ be defined by

$$A_+(x, \lambda) := \{i \in A(x, \lambda) \mid \lambda_i \geq \rho(x, \lambda)\}.$$

The following theorem holds.

THEOREM 2.4. *Let ρ be an identification function for $\mathcal{K}$. If the SMFCQ holds at $\bar{x}$, then there is an $\epsilon > 0$ such that*

$$A_+(x, \lambda) = I_+ \qquad \forall (x, \lambda) \in \mathcal{K} + B_\epsilon.$$

Proof. We first recall that the SMFCQ implies that Λ reduces to a singleton, i.e., $\Lambda = \{\bar{\lambda}\}$. Theorem 2.2 shows that $A_+(x, \lambda) \subseteq I_0$ for all (x, λ) in a certain neighborhood of $(\bar{x}, \bar{\lambda})$. Now, consider an index $i \in I_+$. By continuity, this implies $i \in A_+(x, \lambda)$ in a sufficiently small neighborhood of $(\bar{x}, \bar{\lambda})$. On the other hand, let $i \in I \setminus I_+$. Then, $\bar{\lambda}_i = 0$ and, for all (x, λ) in a sufficiently small neighborhood of $(\bar{x}, \bar{\lambda})$, we have

$$\lambda_i \leq |\lambda_i - \bar{\lambda}_i| \leq \|(x, \lambda) - (\bar{x}, \bar{\lambda})\| = \operatorname{dist}[(x, \lambda), \mathcal{K}] \leq \rho(x, \lambda)/2 < \rho(x, \lambda).$$

This means $i \notin A_+(x, \lambda)$. Thus, $A_+(x, \lambda) = I_+$ for all (x, λ) sufficiently close to $\mathcal{K} = (\bar{x}, \bar{\lambda})$. □

Until now we made reference to the Karush-Kuhn-Tucker system (2.1) which expresses first order necessary optimality conditions for the minimization Problem $(\mathcal{P})$. We showed how the active constraints associated to an isolated stationary point $\bar{x}$ can be identified. However, the fact that the Karush-Kuhn-Tucker system (2.1) derives from an optimization problem plays no role in the theory developed. What we actually proved is the following: Given a solution $(\bar{x}, \bar{\lambda})$ of a system with the structure of system (2.1) and with an isolated x-part, we can identify, in a suitable neighborhood of this solution, those inequalities which hold as equalities at the solution $(\bar{x}, \bar{\lambda})$. Therefore, if we consider the KKT system

$$\begin{aligned} F(x) - \nabla g(x)\lambda &= 0, \\ \lambda &\geq 0, \\ g(x) &\geq 0, \\ \lambda^T g(x) &= 0, \end{aligned} \tag{2.8}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is any continuous function, the theory of this section goes through without any change. This is an important observation, since it allows us to extend the theory developed so far to the identification of active constraints for the *variational inequality problem*:

$$\text{Find } \bar{x} \in X \quad \text{such that} \quad F(\bar{x})^T(x - \bar{x}) \geq 0 \quad \forall x \in X,$$

where $X := \{x \in \mathbb{R}^n | \, g(x) \geq 0\}$, $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and $g : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable. It is well known that, under a standard regularity assumption [19], a necessary condition for $\bar{x} \in \mathbb{R}^n$ to be a solution of the variational inequality problem is that $\bar{\lambda} \in \mathbb{R}^m$ exists such that $(\bar{x}, \bar{\lambda})$ solves system (2.8). Therefore, if we have a sequence $\{(x^k, \lambda^k)\}$ converging to a solution of system (2.8) which has an isolated primal part, we can apply the techniques described in this section in order to identify which of the constraints $g_i(x) \geq 0$ will be active at $\bar{x}$.

3. Defining Identification Functions.

From the previous section we see that the crucial point in the identification of active constraints is the definition of an identification function. In this section we show how it is possible to define such a function for Problem $(\mathcal{P})$.

We consider three cases. In the first, we assume that the functions f and g are analytic. In the second, we require that the functions be LC^1 and that the MFCQ as well as a second order sufficiency condition for optimality be satisfied. Finally, in the third case, the functions are required to be C^2 and the KKT point is assumed to satisfy a regularity condition related to (but weaker than) Robinson's strong regularity [38] and which we call quasi-regularity.

Extensions of these results to the KKT system (2.8) are possible. We shall point out the relevant changes in corresponding remarks.

The cases considered here do not cover all the situations in which an identification function can be defined and computed, but they certainly show that the definition and computation of an identification function is possible in most of the cases of interest.

3.1. The Analytic Case.

Let f and each g_i $(i \in I)$ be *analytic* around a point x. We recall that this means that f and each g_i $(i \in I)$ possess derivatives of all orders and that they agree with their Taylor expansions around x. We say that f and each g_i $(i \in I)$ are analytic on an open set $X \subseteq \mathbb{R}^n$ if they are analytic around each $x \in X$. We shall make use of the following result due to Lojasiewicz, Luo and Pang [30, 32].

THEOREM 3.1. *Let S denote the set of points in $\mathbb{R}^r$ satisfying*

$$s(y) \leq 0, \qquad h(y) = 0,$$

where $s : \mathbb{R}^r \to \mathbb{R}^p$ and $h : \mathbb{R}^r \to \mathbb{R}^t$ are analytic functions defined on an open set $X \subseteq \mathbb{R}^r$. Suppose that $S \neq \emptyset$. Then, for each compact subset $\Omega \subset X$, there exist constants $\tau > 0$ and $\gamma > 0$ such that

$$dist[y, S] \leq \tau(\|[s(y)]_+\| + \|h(y)\|)^\gamma \qquad \forall y \in \Omega. \tag{3.1}$$

Using this result, it is possible to define an identification function for Problem $(\mathcal{P})$.

THEOREM 3.2. *Suppose that f and g are analytic in a neighborhood of a stationary point $\bar{x}$. Then, the function $\rho_1 : \mathbb{R}^n \times \mathbb{R}^m \to [0,\infty)$ defined by*

$$\rho_1(x,\lambda) = \begin{cases} 0 & \text{if } r(x,\lambda) = 0, \\ \frac{-1}{\log(r(x,\lambda))} & \text{if } r(x,\lambda) \in (0, 0.9), \\ \frac{-1}{\log(0.9)} & \text{if } r(x,\lambda) \geq 0.9, \end{cases}$$

where

$$r(x,\lambda) = \|\nabla f(x) - \nabla g(x)\lambda\| + |\lambda^T g(x)| + \|[-\lambda]_+\| + \|[-g(x)]_+\|, \tag{3.2}$$

is an identification function for $\mathcal{K}$.

Proof. It is obvious, by definition, that ρ_1 is a nonnegative function such that $\rho_1(\bar{x},\bar{\lambda}) = 0$ for every $(\bar{x},\bar{\lambda}) \in \mathcal{K}$. Furthermore,

$$\lim_{(x,\lambda)\to(\bar{x},\bar{\lambda})} \rho_1(x,\lambda) = 0 = \rho_1(\bar{x},\bar{\lambda}) \qquad \forall (\bar{x},\bar{\lambda}) \in \mathcal{K},$$

so that ρ_1 is also continuous on $\mathcal{K}$. Hence we only have to check the limit

$$\lim_{\substack{(x,\lambda)\to(\bar{x},\bar{\lambda}) \\ (x,\lambda)\notin\mathcal{K}}} \frac{\rho_1(x,\lambda)}{\operatorname{dist}[(x,\lambda),\mathcal{K}]} = +\infty. \tag{3.3}$$

To this end we recall that, for arbitrary $\tau > 0$ and $\gamma > 0$, the limit

$$\lim_{t\downarrow 0} \frac{-1}{\tau t^\gamma \log t} = +\infty \tag{3.4}$$

holds, see, e.g., [33, p. 328]. We can now apply Theorem 3.1 by considering the system (2.1) which defines KKT points. It is then easy to see that (3.1) yields, for every given compact set $\Omega \subset \mathbb{R}^{n+m}$ containing $(\bar{x},\bar{\lambda})$ in its interior,

$$\operatorname{dist}[(x,\lambda),\mathcal{K}] \leq \tau r(x,\lambda)^\gamma \qquad \forall (x,\lambda) \in \Omega, \tag{3.5}$$

where τ and γ are fixed positive constants. Therefore we can write

$$\lim_{\substack{(x,\lambda)\to(\bar{x},\bar{\lambda}) \\ (x,\lambda)\notin\mathcal{K}}} \frac{\rho_1(x,\lambda)}{\operatorname{dist}[(x,\lambda),\mathcal{K}]} \geq \lim_{\substack{(x,\lambda)\to(\bar{x},\bar{\lambda}) \\ (x,\lambda)\notin\mathcal{K}}} \frac{\rho_1(x,\lambda)}{\tau r(x,\lambda)^\gamma},$$

from which (3.3) follows taking into account (3.4), recalling the definition of ρ_1 and noting that $r(x,\lambda)$ is a continuous function that goes to 0 from the right as (x,λ) tends to $(\bar{x},\bar{\lambda})$. ∎

We stress that Theorem 3.2 holds under the mere assumption that f and g are analytic. In particular, the set of Lagrange multipliers Λ might be unbounded.

REMARK 1. If we want to define an identification function for the solutions of the KKT system (2.8), we only have to substitute the definition of the residual (3.2) by the following one:

$$r(x,\lambda) = \|F(x) - \nabla g(x)\lambda\| + |\lambda^T g(x)| + \|(-\lambda)_+\| + \|(-g(x))_+\|.$$

Obviously, also in this case, we have to assume that F and each g_i $(i \in I)$ are analytic in a neighborhood of the KKT point under consideration.

3.2. The Second Order Condition Case. In this subsection we assume that f and g are LC^1, i.e., that they are differentiable with Lipschitz-continuous derivatives. We denote the *Lagrangian* of problem $(\mathcal{P})$ by

$$L(x,\lambda) := f(x) - \lambda^T g(x)$$

and write $\nabla_x L(x,\lambda)$ for the gradient of L with respect to the x-variables. Furthermore we will make use of the MFCQ and of the following second order sufficient condition for optimality:

Assumption 1. *There is* $\gamma > 0$ *such that, for all* $\bar\lambda \in \Lambda$,

$$h^T H h \geq \gamma \|h\|^2 \qquad \forall h \in W(\bar\lambda),\ \forall H \in \partial_x \nabla_x L(\bar x, \bar\lambda).$$

Here, $W(\bar\lambda)$ *denotes the cone*

$$\{h \in \mathbb{R}^n \mid h^T \nabla g_i(\bar x) \geq 0\ \ (i \in I_0 : \bar\lambda_i = 0),\ h^T \nabla g_i(\bar x) = 0\ \ (i \in I_0 : \bar\lambda_i > 0)\},$$

and $\partial_x \nabla_x L(\bar x, \bar\lambda)$ *denotes Clarke's [8]* generalized Jacobian *with respect to* x *of the gradient* $\nabla_x L$*, calculated at* $(\bar x, \bar\lambda)$.

We remark that, if the functions f and g are twice continuously differentiable and only one multiplier exists, then the previous definition reduces to the classical KKT second order sufficient condition for optimality. Moreover, note that requiring MFCQ implies that the KKT set $\mathcal{K}$ is compact.

Using the MFCQ together with Assumption 1 we will show that the function $\rho_2 : \mathbb{R}^{n+m} \to [0,\infty)$ defined by

$$\rho_2(x,\lambda) := \sqrt{\|\Phi(x,\lambda)\|}$$

is an identification function for $\mathcal{K}$, where the operator $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ is given by

$$\Phi(x,\lambda) := \begin{pmatrix} \nabla_x L(x,\lambda) \\ \min\{g(x),\lambda\} \end{pmatrix}. \tag{3.6}$$

Note that Φ is continuous on $\mathbb{R}^{n+m}$ and that $(x,\lambda) \in \mathcal{K}$ is equivalent to the nonlinear system

$$\Phi(x,\lambda) = 0.$$

To prove that ρ_2 is actually an identification function let us first consider the perturbed nonlinear program

$$\min \quad f(x,t) := f(x) + x^T t_f \quad \text{s.t.} \quad g(x,t) := g(x) + t_g \geq 0, \tag{$\mathcal{P}(t)$}$$

where $t = (t_f, t_g) \in \mathbb{R}^n \times \mathbb{R}^m$ denotes the perturbation parameter. In what follows we will assign to any vector $(y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m$ a particular perturbation vector

$$\tau(y,\mu) = (\tau_f(y,\mu), \tau_g(y,\mu)) \in \mathbb{R}^n \times \mathbb{R}^m.$$

For this purpose we first define the function $\mu^\oplus : \mathbb{R}^{n+m} \to \mathbb{R}^m_+$ componentwise by

$$\mu_i^\oplus(y,\mu) := \begin{cases} \max\{0,\mu_i\} & \text{if } i \in I_\oplus(y,\mu), \\ 0 & \text{if } i \in I \setminus I_\oplus(y,\mu), \end{cases}$$

where, we recall, $I_\oplus(y,\mu) = \{i \in I \mid g_i(y) \le \mu_i\}$. We can now introduce the function $\tau : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ by

$$\begin{aligned} \tau_f(y,\mu) \quad &:= \quad -\nabla_x L(y, \mu^\oplus(y,\mu)), \\ \tau_g(y,\mu)_i \quad &:= \quad \begin{cases} -g_i(y) & \text{if } i \in I_\oplus(y,\mu), \\ -\min\{0, g_i(y)\} & \text{if } i \in I \setminus I_\oplus(y,\mu). \end{cases} \qquad (i \in I) \end{aligned}$$

Using the particular perturbation vector $t = \tau(y,\mu)$, we can prove the following result.

LEMMA 3.3. *Let $(y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m$ be arbitrarily chosen. Then, $(y, \mu^\oplus(y,\mu))$ is a KKT point for problem $(\mathcal{P}(t))$, where $t = \tau(y,\mu)$.*

Proof. The KKT system for the perturbed program $(\mathcal{P}(t))$ reads as follows:

$$\nabla_x L(x,\lambda) + t_f = 0, \tag{3.7}$$

$$\lambda \ge 0, \tag{3.8}$$

$$g(x) + t_g \ge 0, \tag{3.9}$$

$$\lambda^T (g(x) + t_g) = 0. \tag{3.10}$$

Let (y,μ) be arbitrary but fixed. Obviously, since $t = \tau(y,\mu)$, we find that $(x,\lambda) := (y, \mu^\oplus(y,\mu))$ solves (3.7) and (3.8). Now, we will show that $(y, \mu^\oplus(y,\mu))$ also satisfies (3.9) and (3.10). For $i \in I_\oplus(y,\mu)$ the definition of $\tau_g(y,\mu)$ yields $(g(y) + t_g)_i = 0$ so that both (3.9) and (3.10) are fulfilled. If, instead, $i \in I \setminus I_\oplus(y,\mu)$, it follows from the definition of $\mu^\oplus(y,\mu)$ that $\mu_i^\oplus(y,\mu) = 0$ and (3.10) is satisfied. Moreover, the definition of $\tau_g(y,\mu)$ implies

$$(g(y) + t_g)_i = g_i(y) - \min\{0, g_i(y)\} = \max\{0, g_i(y)\} \ge 0 \qquad \forall i \in I \setminus I_\oplus(y,\mu).$$

Thus, (3.9) is also valid for $i \in I \setminus I_\oplus(y,\mu)$. We therefore conclude that $(y, \mu^\oplus(y,\mu))$ is a KKT point of $(\mathcal{P}(t))$ when $t = \tau(y,\mu)$. $\square$

The next lemma is a technical result which will be used in the proof of Theorem 3.6 which, in turn, is the basic ingredient in order to establish the main result of this subsection, Theorem 3.7 below.

LEMMA 3.4.

(a) *It holds that*

$$\|\mu - \mu^\oplus(y,\mu)\| \le \|\min\{g(y),\mu\}\| \le \|\Phi(y,\mu)\| \qquad \forall (y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m.$$

(b) *If the MFCQ is satisfied then $\kappa > 0$ exists such that*

$$\|\tau(y,\mu)\| \le \kappa \|\Phi(y,\mu)\| \qquad \forall (y,\mu) \in \mathcal{K} + B_1.$$

Proof. Let us consider any $(y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m$. We easily see that

$$\min\{g_i(y), \mu_i\} = \begin{cases} g_i(y) \quad \le \quad \mu_i & \text{if } i \in I_\oplus(y,\mu) \\ \mu_i \quad \le \quad g_i(y) & \text{if } i \in I \setminus I_\oplus(y,\mu) \end{cases}.$$

Moreover, this and the definitions of the functions $\mu^\oplus$ and τ_g yield, for $i \in I_\oplus$,

$$\begin{aligned} |\mu_i - \mu_i^\oplus(y,\mu)| \quad &= \quad |\min\{0,\mu_i\}| \quad \le \quad |\min\{g_i(y),\mu_i\}|, \\ |\tau_g(y,\mu)_i| \quad &= \quad |g_i(y)| \quad = \quad |\min\{g_i(y),\mu_i\}|. \end{aligned}$$

Similarly, for $i \in I \setminus I_\oplus(y,\mu)$, we get

$$\begin{array}{lclcl} |\mu_i - \mu_i^\oplus(y,\mu)| & = & |\mu_i| & = & |\min\{g_i(y),\mu_i\}|, \\ |\tau_g(y,\mu)_i| & = & |\min\{0,g_i(y)\}| & \leq & |\min\{g_i(y),\mu_i\}|. \end{array}$$

Thus, property (a) and

$$\|\tau_g(y,\mu)\| \leq \|\min\{g(y),\mu\}\| \leq \|\Phi(y,\mu)\| \qquad \forall (y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m \tag{3.11}$$

follow. To prove (b) let $(y,\mu) \in \mathcal{K} + B_1$. Since, due to the MFCQ, $\mathcal{K}+B_1$ is bounded the LC^1 property of f and g implies that the function $\|\nabla_x L\|$ is globally Lipschitz-continuous on $\mathcal{K}+B_1$ with some modulus $\kappa_0 > 0$. Using property (a) and (3.11), we therefore obtain

$$\begin{array}{rcl} \|\tau(y,\mu)\| & \leq & \|\tau_f(y,\mu)\| + \|\tau_g(y,\mu)\| \\ & \leq & \|\nabla_x L(y,\mu)\| + \kappa_0\|\mu - \mu^\oplus(y,\mu)\| + \|\tau_g(y,\mu)\| \\ & \leq & \kappa\|\Phi(y,\mu)\| \end{array}$$

with $\kappa := \kappa_0 + 2$. ☐

The next result can easily be derived from Theorem 4.5 b) and formula (3.2 f) in Klatte [23]. If the functions f and g of the program $(\mathcal{P})$ are twice continuously differentiable it can also be obtained from a corresponding result in Robinson [39, Corollary 4.3]. We further note that Assumption 1 can be weakened by using generalized directional derivatives, see [23] for more details and references.

THEOREM 3.5. *Let the MFCQ and Assumption 1 be satisfied. Then, there are $\delta > 0$, $\eta > 0$ and $c > 0$ such that*

$$dist[(\bar{x}(t),\bar{\lambda}(t)),\mathcal{K}] \leq c\|t\|$$

for every $t \in B_\delta$ and for every KKT point $(\bar{x}(t),\bar{\lambda}(t))$ of problem $(\mathcal{P}(t))$ for which $\bar{x}(t) \in \{\bar{x}\} + B_\eta$.

Putting together the last three results, we can prove the following theorem.

THEOREM 3.6. *Let the MFCQ and Assumption 1 be satisfied. Then, there are $\epsilon > 0$, $\kappa_1 > 0$ and $\kappa_2 > 0$ such that*

$$\kappa_1 dist[(y,\mu),\mathcal{K}] \leq \|\Phi(y,\mu)\| \leq \kappa_2 dist[(y,\mu),\mathcal{K}] \qquad \forall (y,\mu) \in \mathcal{K} + B_\epsilon.$$

Proof. Let us consider any $(y,\mu) \in \mathbb{R}^n \times \mathbb{R}^m$ and let $z_1 \in \mathcal{K}$ and $z_2 \in \mathcal{K}$ be the projections of (y,μ) and $(y,\mu^\oplus(y,\mu))$, respectively, on the closed convex set $\mathcal{K}$. Then, using the triangle inequality, we get

$$\begin{array}{rcl} \mathrm{dist}[(y,\mu),\mathcal{K}] & = & \|z_1 - (y,\mu)\| \\ & \leq & \|z_2 - (y,\mu)\| \\ & \leq & \|z_2 - (y,\mu^\oplus(y,\mu))\| + \|(y,\mu) - (y,\mu^\oplus(y,\mu))\| \\ & = & \mathrm{dist}[(y,\mu^\oplus(y,\mu)),\mathcal{K}] + \|\mu - \mu^\oplus(y,\mu)\| \end{array} \tag{3.12}$$

Now we will provide an upper bound for $\mathrm{dist}[(y,\mu^\oplus(y,\mu)),\mathcal{K}]$. Taking into account Lemma 3.4 (b) and that $\|\Phi\|$ is a continuous function with $\Phi(y,\mu) = 0$ for all $(y,\mu) \in \mathcal{K}$, we have that, for δ from Theorem 3.5, we can find $\bar{\epsilon} > 0$ such that, if $(y,\mu) \in \mathcal{K}+B_{\bar{\epsilon}}$

then $\|\tau(y,\mu)\| \le \kappa\|\Phi(y,\mu)\| \le \delta$. Therefore, since $\epsilon \le \eta$ (with η from Theorem 3.5) can be assumed without loss of generality, Theorem 3.5 together with Lemma 3.3 yield

$$\mathrm{dist}[(y,\mu^{\oplus}(y,\mu)),\mathcal{K}] \le c\|\tau(y,\mu)\| \qquad \forall (y,\mu) \in \mathcal{K} + B_\epsilon.$$

Using this, (3.12) and Lemma 3.4, we obtain

$$\mathrm{dist}[(y,\mu),\mathcal{K}] \le c\|\tau(y,\mu)\| + \|\mu - \mu^{\oplus}(y,\mu)\| \le (c\kappa+1)\|\Phi(y,\mu)\| \quad \forall (y,\mu) \in \mathcal{K} + B_\epsilon,$$

i.e., the left inequality in the theorem is satisfied with $\kappa_1 := 1/(c\kappa+1)$. The right inequality can easily be obtained by taking into account that $\mathcal{K}$ is compact and convex and that $\|\Phi\|$ is locally Lipschitz-continuous. Therefore, $\kappa_2 > 0$ exists such that

$$\|\Phi(y,\mu)\| = \|\Phi(y,\mu) - \Phi(z_1)\| \le \kappa_2\|(y,\mu) - z_1\| = \kappa_2 \mathrm{dist}[(y,\mu),\mathcal{K}] \quad \forall (y,\mu) \in \mathcal{K} + B_\epsilon,$$

where (as above) z_1 denotes the projection of (y,μ) onto $\mathcal{K}$. □

THEOREM 3.7. *Let the MFCQ and Assumption 1 be satisfied. Then ρ_2 is an identification function for $\mathcal{K}$.*

Proof. Taking the properties of the operator Φ into account we easily see that ρ_2 is nonnegative and continuous on $\mathbb{R}^{n+m}$ and that $\rho_2(x,\lambda) = 0$ for all $(x,\lambda) \in \mathcal{K}$, so that properties (a) and (b) of Definition 2.1 are satisfied. Finally, property (c) immediately follows from Theorem 3.6. □

If, instead of the upper Lipschitz-continuity as stated in Theorem 3.5, the multifunction $t \mapsto \mathcal{K}(t)$ is upper Hölder-continuous at $t = 0$ with a known rate $\nu \in (0,1]$, that is, if, for some $\delta > 0$, $\eta > 0$ and $c > 0$,

$$\mathrm{dist}\,[(\bar{x}(t),\bar{\lambda}(t)),\mathcal{K}] \le c\|t\|^{\nu}$$

for every $t \in B_\delta$ and for every KKT point $(\bar{x}(t),\bar{\lambda}(t))$ of Problem $(\mathcal{P}(t))$ for which $\bar{x}(t) \in \{\bar{x}\}+B_\eta$, then the technique presented in this subsection can easily be extended if we define $\rho_2 : \mathbb{R}^n \times \mathbb{R}^m \to [0,\infty)$ as

$$\rho_2(x,\lambda) := \|\Phi(x,\lambda)\|^{\nu/2}.$$

In particular, Theorem 3.7 remains valid for this ρ_2 if Assumption 1 is replaced by the upper Hölder-continuity.

An interesting case in which it is possible to prove, under an assumption weaker than Assumption 1, the upper Hölder-continuity at $t = 0$ of the multifunction $t \mapsto \mathcal{K}(t)$ is the case of convex problems. Assume that f is convex and each g_i $(i \in I)$ is concave, that the MFCQ holds and that the following *growth condition* holds (in place of Assumption 1): positive $\bar{\eta}$ and $\bar{c}$ exist such that

$$f(x) \ge f(\bar{x}) + \bar{c}\|x - \bar{x}\|^2, \qquad \text{for all feasible } x \text{ in } \{\bar{x}\} + B_{\bar{\eta}}.$$

Under these assumptions and using the results in [24], it is possible to show (we omit the details) that $\delta > 0$, $\eta > 0$ and $c > 0$ exist such that

$$\mathrm{dist}\,[(\bar{x}(t),\bar{\lambda}(t)),\mathcal{K}] \le c\sqrt{\|t\|}$$

for every $t \in B_\delta$ and for every KKT point $(\bar{x}(t),\bar{\lambda}(t))$ of Problem $(\mathcal{P}(t))$ for which $\bar{x}(t) \in \{\bar{x}\} + B_\eta$. It may be interesting to note that the growth condition holds, in particular, if Assumption 1 is fulfilled.

REMARK 2. The extension of the results of this section to general KKT systems is not straightforward, since the sensitivity analysis of perturbed KKT systems requires, to date, stronger assumptions. The key point is to establish a result analogous to Theorem 3.5. Once this has been done, we can easily prove theorems analogous to Theorem 3.7 by substituting F to ∇f in every relevant formula. As an example of the kind of the results that can be obtained we cite the following one. Suppose that F is C^1 and g is C^2. Assume also that the SMFCQ holds at $\bar{x}$ along with Assumption 1. Then, according to [18, Corollary 8 (c)], Theorem 3.5 holds and therefore ρ_2 is a regular identification function for the KKT system (2.8).

3.3. The Quasi-Regular Case. In this subsection we assume that the functions f and g are C^2. We shall introduce a condition which we call *quasi-regularity.* As will be clear later, this quasi-regularity is related to , but weaker than Robinson's *strong regularity* [38]. In order to motivate the definition of a quasi-regular KKT point we will first recall a condition which is equivalent to the notion of a strongly regular KKT point. To this end we shall use the index set $I_{00} := I_0 \setminus I_+$ of all those indices for which the strict complementarity condition does not hold at the KKT point $(\bar{x}, \bar{\lambda})$. For any $J \subseteq I_{00}$ (empty set included) introduce the matrix

$$M(J) := \begin{pmatrix} \nabla^2_{xx} L & \nabla g_+ & \nabla g_J \\ -\nabla g_+^T & 0 & 0 \\ -\nabla g_J^T & 0 & 0 \end{pmatrix},$$

where $\nabla^2_{xx} L$, ∇g_+ and ∇g_J are abbreviations for the matrices $\nabla^2_{xx} L(\bar{x}, \bar{\lambda})$, $\nabla g_{I_+}(\bar{x})$ and $\nabla g_J(\bar{x})$, respectively. The following result is due to Kojima [26].

THEOREM 3.8. *The following statements are equivalent:*

(a) $(\bar{x}, \bar{\lambda})$ *is a strongly regular KKT point.*

(b) For any $J \subseteq I_{00}$ *(empty set included), the determinants of the matrices* $M(J)$ *all have the same nonzero sign.*

Motivated by point (b) in Theorem 3.8, we introduce the following definition.

DEFINITION 3.9. *The KKT point* $(\bar{x}, \bar{\lambda})$ *is a* quasi-regular *point if the matrices* $M(J)$ *are nonsingular for every* $J \subseteq I_{00}$ *(empty set included).*

Note that, in view of Theorem 3.8, quasi-regularity is implied by Robinson's strong regularity condition, but the converse is not true. In fact, consider the following example:

$$\begin{aligned} &\min x_1^2 + x_2^2 + 4x_1x_2 \\ &\text{s.t.} x_1 \geq 0, \\ &\quad\ x_2 \geq 0. \end{aligned}$$

It is easy to check that $\bar{x} = (0,0)$ is a global minimizer and that the Lagrange multipliers of the two constraints are both zero, so that $I_0 = I_{00} = \{1,2\}$, while $I_+ = \emptyset$. Furthermore, $detM(\emptyset) < 0$, while, for $J \in \{\{1\},\{2\},\{1,2\}\}$, $detM(J) > 0$. Therefore $(\bar{x}, 0, 0)$ is a quasi-regular KKT point, but not a strongly regular one. Note that in this example the KKT point is an isolated KKT point. This is not a coincidence. In fact, we shall show in this section that quasi-regularity of a KKT point implies its local uniqueness. It is also worth pointing out that quasi-regularity implies the linear independence of the active constraints. This easily follows from the fact that $M(I_{00})$ is nonsingular.

As in Subsection 3.2 we make use of the operator $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined in (3.6) which, due to the differentiability assumptions, is locally Lipschitz-continuous.

Hence, by Rademacher's Theorem, Φ is differentiable almost everywhere. Denote by D_Φ the set of points where Φ is differentiable. Then we can define the *B-subdifferential* (see, e.g., [36]) of Φ at (x,λ) as

$$\partial_B\Phi(x,\lambda) := \{H \in \mathbb{R}^{(n+m)\times(n+m)} \mid \exists\{(x^k,\lambda^k)\} \subset D_\Phi : \\ (x^k,\lambda^k) \to (x,\lambda),\ \nabla\Phi(x^k,\lambda^k)^T \to H\}$$

Note that the B-subdifferential is a subset of Clarke's generalized Jacobian [8, 36]. The next lemma illustrates the structure of the B-subdifferential of Φ. Before stating this lemma, however, we introduce three index sets:

$$\begin{aligned} \alpha(x,\lambda) &:= \{i \in I \mid g_i(x) < \lambda_i\}, \\ \beta(x,\lambda) &:= \{i \in I \mid g_i(x) = \lambda_i\}, \\ \gamma(x,\lambda) &:= \{i \in I \mid g_i(x) > \lambda_i\}. \end{aligned}$$

LEMMA 3.10. *Let $(x,\lambda) \in \mathbb{R}^{n+m}$ be arbitrary. Then*

$$\partial_B\Phi(x,\lambda)^T \subseteq \begin{pmatrix} \nabla^2_{xx}L(x,\lambda) & \nabla g(x)D_a(x,\lambda) \\ -\nabla g(x)^T & D_b(x,\lambda) \end{pmatrix},$$

where

$$\begin{aligned} D_a(x,\lambda) &:= diag\,(a_1(x,\lambda),\dots,a_m(x,\lambda)), \\ D_b(x,\lambda) &:= diag\,(b_1(x,\lambda),\dots,b_m(x,\lambda)) \end{aligned}$$

are diagonal matrices with

$$a_i(x,\lambda) = \begin{cases} 1 & \text{if } i \in \alpha(x,\lambda), \\ 0 \text{ or } 1 & \text{if } i \in \beta(x,\lambda), \\ 0 & \text{if } i \in \gamma(x,\lambda), \end{cases}$$

and $D_b(x,\lambda) = I - D_a(x,\lambda)$.

Proof. This follows immediately from the definition of the operator Φ. □

We are now in the position to prove the following result.

LEMMA 3.11. *Let $(\bar{x},\bar{\lambda}) \in \mathbb{R}^{n+m}$ be a quasi-regular KKT point. Then all matrices $H \in \partial_B\Phi(\bar{x},\bar{\lambda})$ are nonsingular.*

Proof. Let $H \in \partial_B\Phi(\bar{x},\bar{\lambda})^T$. In view of Lemma 3.10, there exists an index set $J \subseteq \beta(\bar{x},\bar{\lambda})$ such that

$$H = \begin{pmatrix} \nabla^2_{xx}L & \nabla g_\alpha & \nabla g_J & 0 & 0 \\ -\nabla g_\alpha^T & 0 & 0 & 0 & 0 \\ -\nabla g_J^T & 0 & 0 & 0 & 0 \\ -\nabla g_{\bar{J}}^T & 0 & 0 & I_{\bar{J}} & 0 \\ -\nabla g_\gamma^T & 0 & 0 & 0 & I_\gamma \end{pmatrix}$$

where $\bar{J} = \beta(\bar{x},\bar{\lambda}) \setminus J$ denotes the complement of J in the set $\beta(\bar{x},\bar{\lambda})$. Obviously, this matrix is nonsingular if and only if the matrix

$$\begin{pmatrix} \nabla^2_{xx}L & \nabla g_\alpha & \nabla g_J \\ -\nabla g_\alpha^T & 0 & 0 \\ -\nabla g_J^T & 0 & 0 \end{pmatrix}$$

is nonsingular. In turn, this matrix is nonsingular if and only if the matrix $M(J)$ is nonsingular. Hence the thesis follows immediately from Definition 3.9. □

We are now able to prove the main result of this subsection. For this purpose recall that $\rho_2(x,\lambda) = \sqrt{\|\Phi(x,\lambda)\|}$ (see Subsection 3.2).

THEOREM 3.12. *Let* $(\bar{x},\bar{\lambda}) \in \mathbb{R}^{n+m}$ *be a quasi-regular KKT point of problem* $(\mathcal{P})$*. Then,*

(a) $(\bar{x},\bar{\lambda})$ *is an isolated KKT point,*

(b) the function ρ_2 *is an identification function for* $\mathcal{K} = \{(\bar{x},\bar{\lambda})\}$*.*

Proof. As already shown in the proof of Theorem 3.7 the function ρ_2 has the properties (a) and (b) of Definition 2.1. Furthermore, since f and g have locally Lipschitz-continuous gradients and the min operator is semismooth (see [34, 37] for the definition of semismoothness and [34] for the proof that the min operator is semismooth) it follows that also Φ, which is the composite of semismooth functions, is semismooth [34, 37]. Hence it follows from Lemma 3.11 and [35, Proposition 3] that there exists a constant $c > 0$ such that

$$\|\Phi(x,\lambda)\| \geq c\|(x,\lambda) - (\bar{x},\bar{\lambda})\| = c\,\mathrm{dist}[(x,\lambda),\mathcal{K}] \tag{3.13}$$

for all (x,λ) in a neighborhood of $(\bar{x},\bar{\lambda})$. Therefore, one can easily see that ρ_2 also has property (c) of Definition 2.1, i.e., ρ_2 is an identification function for $\mathcal{K}$. Finally, since $\Phi(x,\lambda) = 0$ if and only if (x,λ) is a KKT point, part (a) of the theorem follows from (3.13). □

REMARK 3. In the case of the KKT system (2.8) everything goes through. It is sufficient to assume that F is continuously differentiable and to substitute everywhere the gradient $\nabla_x L(x,\lambda)$ by the function $F(x) - \nabla g(x)\lambda$. Also in this case the definition of quasi-regularity is related to and weaker than that of a strongly regular KKT point since Theorem 3.8 carries over to the KKT system (2.8), see Liu [29, Lemma 3.4]. Actually, the case of KKT systems of variational inequalities is probably the main case in which quasi-regularity can be applied. In fact, it is not difficult to see that, if strict complementarity holds and $\bar{x}$ is a local minimum point of Problem $(\mathcal{P})$, quasi-regularity implies the conditions of the previous subsection. However, these conditions and quasi-regularity are fairly distinct if one considers variational inequalities. For example, it can easily be checked that, given the variational inequality defined by the function $F(x) = (x_1 + x_2^2, -x_2)^T$ and the set $X = \{x \in \mathbb{R}^2 | \, x_2 \geq 0\}$, the point $(0,0)^T$ is a quasi-regular solution but does not satisfy the conditions stated in Remark 2 of the previous subsection.

4. Numerical Examples. In this section we illustrate the identification technique on three nonlinear optimization problems. Our aim here is merely to give the reader a feel of the potentialities of the new technique. A detailed study of its numerical behavior is out of the scope of this paper.

We consider three test problems from the Hock and Schittkowski collection [20]. The first is problem 113 and at the solution both the linear independence constraint qualification and the strict complementarity condition are satisfied. The second problem is a modification of problem 46 and, while the linear independence constraint qualification is satisfied at the solution, the multipliers are all zero. Finally, we consider a modification of problem 43 whose multiplier set Λ is not a singleton.

For these test problems we applied the identification technique for both identifications functions ρ_1 and ρ_2 introduced in Section 3. To this end random points (x,λ) at different fixed distances from the set $\mathcal{K}$ were generated. More in detail, for each

$\varepsilon \in \{10, 1, 10^{-1}, 10^{-2}, 10^{-3}\}$, we generated 100 random vectors (x, λ) on the boundary of the set

$$\mathcal{K} + B_\varepsilon^\infty = \{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, \exists \bar{\lambda} \in \Lambda \,:\, \|(x,\lambda) - (\bar{x}, \bar{\lambda})\|_\infty < \varepsilon\}.$$

For each of these random vectors we compared our approximate active sets $A(x, \lambda)$ with the exact active set I_0. For each of the constraints, for the different values of ε and for both identification functions ρ_1 and ρ_2, we report the sum of the correctly identified constraints over all 100 randomly generated vectors (x, λ), see the tables below. The last column of each table contains the total number of correctly identified constraints over all constraints.

Example 1. This is problem 113 from [20]. It is a convex optimization problem with $n = 10$ variables and $m = 8$ inequality constraints, five of them being nonlinear. The solution is given by

$$\bar{x} \approx (2.17, 2.36, 8.77, 5.10, 0.99, 1.43, 1.32, 9.83, 8.28, 8.38)^T,$$

and the corresponding optimal Lagrange multiplier is unique and given by

$$\bar{\lambda} \approx (1.72, 0.48, 1.38, 0.02, 0.31, 0, 0.29, 0)^T.$$

The solution satisfies the strict complementarity condition; however, since the fourth constraint is active and $\bar{\lambda}_4 \approx 0.02$, the solution is relatively close to being degenerate. Our results are summarized in Table 4.1.

Table 4.1: Numerical results for Example 1

ε	ρ	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_1 - g_8
$\varepsilon = 10$	ρ_1	54	54	57	89	90	2	78	16	440
	ρ_2	65	60	71	90	93	0	84	12	475
$\varepsilon = 1$	ρ_1	90	76	94	68	75	22	83	100	608
	ρ_2	81	68	86	64	67	36	75	100	577
$\varepsilon = 0.1$	ρ_1	100	100	100	100	100	0	100	100	700
	ρ_2	100	94	100	76	90	100	99	100	759
$\varepsilon = 0.01$	ρ_1	100	100	100	100	100	82	100	100	782
	ρ_2	100	100	100	100	100	100	100	100	800
$\varepsilon = 0.001$	ρ_1	100	100	100	100	100	100	100	100	800
	ρ_2	100	100	100	100	100	100	100	100	800

Example 2. This example is a modification of problem 46 from [20]. Problem 46 has two equalities constraints which have zero multipliers at the solution. We converted the equalities to inequalities and added the constraint $x_2 \leq 1$ in order to maintain the uniqueness of the solution considered. Thus we have $n = 5$ variables and $m = 3$ inequality constraints. The objective function is given by

$$f(x) := (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6,$$

and the constraints are

$$g_1(x) := x_1^2 x_4 + \sin(x_4 - x_5) - 1 \geq 0,$$
$$g_2(x) := x_2 + x_3^4 x_4^2 - 2 \geq 0,$$
$$g_3(x) := 1 - x_2 \geq 0.$$

The solution is

$$\bar{x} := (1, 1, 1, 1, 1)^T$$

and the corresponding multiplier is

$$\bar{\lambda} := (0, 0, 0)^T.$$

Since all inequality constraints are active at the solution $\bar{x}$, $(\bar{x}, \bar{\lambda})$ is totally degenerate. We report our results in Table 4.2.

Table 4.2: Numerical results for Example 2

ε	ρ	g_1	g_2	g_3	g_1 - g_3
$\varepsilon = 10$	ρ_1	52	8	92	152
	ρ_2	85	18	100	203
$\varepsilon = 1$	ρ_1	100	82	100	282
	ρ_2	88	73	100	261
$\varepsilon = 0.1$	ρ_1	100	99	100	299
	ρ_2	100	97	100	297
$\varepsilon = 0.01$	ρ_1	100	100	100	300
	ρ_2	100	100	100	300
$\varepsilon = 0.001$	ρ_1	100	100	100	300
	ρ_2	100	100	100	300

EXAMPLE 3. This example is a modification of problem 43 from [20]. It has $n = 4$ variables and $m = 4$ inequality constraints. Its objective function is

$$f(x) := x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4,$$

and its constraints are

$$\begin{aligned} g_1(x) &:= -x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 + 8 \geq 0, \\ g_2(x) &:= -x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 + 10 \geq 0, \\ g_3(x) &:= -2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 + 5 \geq 0, \\ g_4(x) &:= x_2^3 + 2x_1^2 + x_4^2 + x_1 - 3x_2 - x_3 + 4x_4 + 7 \geq 0, \end{aligned}$$

i.e., we added the fourth constraint to problem 43 from [20]. The solution of this problem is

$$\bar{x} = (0, 1, 2, -1)^T.$$

The constraints g_1, g_3 and g_4 are active at the solution, and

$$\nabla g_4(\bar{x}) = \nabla g_1(\bar{x}) - \nabla g_3(\bar{x})$$

so that the linear independence constraint qualification is violated. However, the corresponding set of Lagrange multipliers, given by

$$\Lambda := \{\bar{\lambda}(r) := (3 - r, 0, r, r - 2)^T \mid r \in [2, 3]\},$$

is bounded, so that the Mangasarian-Fromovitz constraint qualification is satisfied. Furthermore, if $r \in \{2, 3\}$, then strict complementarity is violated.

To test this problem, the random points (x, λ) on the boundary of $\mathcal{K} + B_\varepsilon^\infty$ were generated as follows. First the x-part was randomly generated such that $\|x - \bar{x}\|_\infty = \varepsilon$. To obtain the λ-part we took a random number $r \in [2, 3]$ and then generated the vector λ randomly such that $\|\lambda - \bar{\lambda}(r)\|_\infty = \varepsilon$. It is obvious that every point (x, λ) generated in this way lies on the boundary of $\mathcal{K} + B_\varepsilon^\infty$. In Table 4.3 we summarize the results obtained for this example.

Table 4.3: Numerical results for Example 3

ε	ρ	g_1	g_2	g_3	g_4	g_1 - g_4
$\varepsilon = 10$	ρ_1	100	0	100	26	226
	ρ_2	100	0	100	27	227
$\varepsilon = 1$	ρ_1	100	0	100	100	300
	ρ_2	89	18	96	65	268
$\varepsilon = 0.1$	ρ_1	100	0	100	100	300
	ρ_2	100	36	100	100	336
$\varepsilon = 0.01$	ρ_1	100	97	100	100	397
	ρ_2	100	100	100	100	400
$\varepsilon = 0.001$	ρ_1	100	100	100	100	400
	ρ_2	100	100	100	100	400

We think these three examples suggest that the identification technique is viable in practice even if we are well aware that no firm conclusion can be drawn on the basis of these few tests.

It is also important to point out that if ρ is an identification function, also any positive multiple of ρ is an identification function; in practice an appropriate scaling of the identification functions might be crucial for a good performance of the identification technique. Finally we note that if one wants to employ the identification technique in combination with a specific solution algorithm, one should take into account that sequences generated by specific algorithms may have additional properties which should be exploited to enhance the identification process.

5. Final Remarks. In this paper we introduced a technique to accurately identify active constraints in inequality constrained optimization and variational inequality problems. The most remarkable features of the new identification technique are on the one hand that it identifies all active constraints even if strict complementarity does not hold and, on the other hand, that, as far as we are aware of, it is the first identification technique applicable to nonlinear variational inequalities. Furthermore, as discussed in the introduction, it also enjoys several other favorable characteristics. In particular, the identification technique can be used in combination with any algorithm for the solution of inequality constrained optimization or variational inequality problems.

We believe that the techniques introduced in this paper can be useful in many cases, especially in the theoretical analysis and design of optimization methods.

From a practical point of view, the following questions may be of interest:

(a) How large is the region where exact identification occurs?
(b) Can we build identification functions which are scale invariant?
(c) Can we relax the assumption that $\bar{x}$ is an isolated stationary point and still obtain useful results?

It is difficult to answer to these questions at the level of generality adopted in this paper. We think that an answer can come from practical experiments and from an analysis of structured classes of problems, e.g., linear or quadratic problems, box or linearly constrained problems etc.

From a more theoretical point of view we would like to mention that the identification technique introduced in this paper turned out to be an essential tool in the development of the first algorithm for nonlinearly inequality constrained problems for which convergence to points satisfying the second order necessary condition for optimality can be established, see [14]. Moreover, the identification technique is one basic ingredient for the algorithm suggested in [22] which is the first QP-free method for the solution of variational inequality problems which has a global and superlinear convergence and which generates (in some sense) only feasible iterates. Finally, let us mention that the new identification technique has been advocated in [43] to accommodate a theoretical assumption needed to establish the superlinear convergence of an SQP-type method even when the linear independence of the active constraints is not satisfied at a solution.

Acknowledgments. We would like to thank Professor D. Klatte for helpful discussions on the stability of KKT systems.

REFERENCES

[1] R. Bellman: *Introduction to Matrix Analysis.* McGraw-Hill, New York, 1970.

[2] D.P. Bertsekas: *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[3] J.F. Bonnans: *Rates of convergence of Newton type methods for variational inequalities and nonlinear programming.* Applied Mathematics and Optimization 29, 1994, pp. 161–186.

[4] J.V. Burke: *On the identification of active constraints II: The nonconvex case.* SIAM Journal on Numerical Analysis 27, 1990, pp. 1081–1102.

[5] J.V. Burke and J.J. Moré: *On the identification of active constraints.* SIAM Journal on Numerical Analysis 25, 1988, pp. 1197–1211.

[6] J.V. Burke and J.J. Moré: *Exposing constraints.* SIAM Journal on Optimization 4, 1994, pp. 573–595.

[7] J.V. Burke, J.J. Moré and G. Toraldo: *Convergence properties of trust region methods for linear and convex constraints.* Mathematical Programming 47, 1990, pp. 305–336.

[8] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley and Sons, New York, 1983 (reprinted by SIAM, Philadelphia, 1990).

[9] A.R. Conn, N.I.M. Gould and P.L. Toint: *Global convergence for a class of trust region algorithms for optimization problems with simple bounds.* SIAM Journal on Numerical Analysis 25, 1988, pp. 433–460.

[10] A.S. El-Bakry, R.A. Tapia and Y. Zhang: *A study of indicators for identifying zero variables in interior-point methods.* SIAM Review 36, 1994, pp. 45–72.

[11] A.S. El-Bakry, R.A. Tapia and Y. Zhang: *On the convergence rate of Newton interior-point methods in the absence of strict complementarity.* Computational Optimization and Applications 6, 1996, pp. 157–167.

[12] F. Facchinei and S. Lucidi: *A class of methods for optimization problems with simple bounds.* Technical Report, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Rome, Italy, 1992 (revised 1995).

[13] F. Facchinei and S. Lucidi: *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems.* Journal of Optimization Theory and Applications 85, 1995, pp. 265–289.

[14] F. Facchinei and S. Lucidi: *Convergence to second order stationary points in inequality constrained optimization.* DIS Working Paper 32-96, Università di Roma "La Sapienza", Roma, Italy, 1996. To appear in Mathematics of Operations Research.

[15] R. Fletcher: *Practical Methods of Optimization.* John Wiley and Sons, New York, 1987.

[16] J. Gauvin: *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming.* Mathematical Programming 12, 1977, pp. 136–138.

[17] P.E. Gill, W. Murray and M.H. Wright: *Practical Optimization.* Academic Press, London, 1981.

[18] M.S. Gowda and J.-S. Pang: *Stability analysis of variational inequalities and nonlinear complementarity problems, via the mixed linear complementarity problem and degree theory.* Mathematics of Operations Research 19, 1994, pp. 831–879.

[19] P.T. Harker and J.-S. Pang: *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications.* Mathematical Programming 48, 1990, pp. 161–220.

[20] W. Hock and K. Schittkowski: *Test examples for nonlinear programming codes.* Lectures Notes in Economics and Mathematical Systems 187, Springer-Verlag, Berlin, 1981.

[21] J. Ji and F.A. Potra: *Tapia indicators and finite termination of infeasible-interior-point methods for degenerate LCP.* Lectures in Applied Mathematics 32, 1996, pp. 443–454.

[22] C. Kanzow and H.-D. Qi: *A QP-free constrained Newton-type method for variational inequality problems.* Preprint 121, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, 1997.

[23] D. Klatte: *Nonlinear optimization problems under data perturbations.* In: W. Krabs and J. Zowe (eds.): *Modern Methods of Optimization.* Springer-Verlag, Berlin, 1992, pp. 204–235.

[24] D. Klatte: *On quantitative stability for $C^{1,1}$ programs.* In: R. Durier and C. Michelot (eds.): *Recent Developments in Optimization.* Springer-Verlag, Berlin, 1995, pp. 215–230.

[25] H. Kleinmichel, C. Richter and K. Schönefeld: *On a class of hybrid methods for smooth constrained optimization.* Journal of Optimization Theory and Applications 73, 1992, pp. 465–499.

[26] M. Kojima: *Strongly stable stationary solutions in nonlinear programs.* In: S.M. Robinson (ed.): *Analysis and Computation of Fixed Points.* Academic Press, New York, 1979, pp. 93–138.

[27] J. Kyparisis: *On uniqueness of Kuhn Tucker multipliers in nonlinear programming.* Mathematical Programming 32, 1985, pp. 242–246.

[28] M. Lescrenier: *Convergence of trust region algorithms for optimization with bounds when strict complementarity does not hold.* SIAM Journal on Numerical Analysis 28, 1991, pp. 476–495.

[29] J. Liu: *Strong stability in variational inequalities.* SIAM Journal on Control and Optimization 33, 1995, pp. 725–749.

[30] M.S. Lojasiewicz: *Sur le problème de la division.* Studia Mathematica 18, 1959, pp. 87–136.

[31] S. Lucidi: *New results on a continuously differentiable exact penalty function.* SIAM Journal on Optimization 2, 1992, pp. 558–574.

[32] Z.-Q. Luo and J.-S. Pang: *Error bounds for analytic systems and their applications.* Mathematical Programming 67, 1994, pp. 1–28.

[33] J. Marsden and A. Weinstein: *Calculus I.* Springer-Verlag, New York, 1985.

[34] R. Mifflin: *Semismooth and semiconvex functions in constrained optimization.* SIAM Journal on Control and Optimization 15, 1977, pp. 957–972.

[35] J.-S. Pang and L. Qi: *Nonsmooth equations: motivation and algorithms.* SIAM Journal on Optimization 3, 1993, pp. 443–465.

[36] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[37] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–368.

[38] S.M. Robinson: *Strongly regular generalized equations.* Mathematics of Operations Research 5, 1980, pp. 43–62.

[39] S.M. Robinson: *Generalized equations and their solution, part II: Applications to nonlinear programming.* Mathematical Programming Study 19, 1982, pp. 200–221.

[40] K. Schönefeld: *Hybrid optimization methods without strict complementary slackness conditions.* Proceedings of the "International Conference on Mathematical Optimization — Theory and Applications", Eisenach, Germany, 1986, pp. 137–140.

[41] S.J. Wright: *Convergence of SQP like methods for constrained optimization.* SIAM Journal on Control and Optimization 27, 1989, pp. 13–26.

[42] S.J. Wright: *Identifiable surfaces in constrained optimization.* SIAM Journal on Control and Optimization 31, 1993, pp. 1063–1079.

[43] S.J. Wright: *Superlinear convergence of a stabilized SQP method to a degenerate solution.* Preprint ANL/MCS-P643-0297. Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 1997.