Mathematical Programming
(accepted for publication)

KARUSH–KUHN–TUCKER SYSTEMS: REGULARITY CONDITIONS, ERROR BOUNDS AND A CLASS OF NEWTON-TYPE METHODS*

A. F. Izmailov† and M. V. Solodov‡

October 29, 2001 (Revised May 2, 2002)

ABSTRACT

We consider optimality systems of Karush-Kuhn-Tucker (KKT) type, which arise, for example, as primal-dual conditions characterizing solutions of optimization problems or variational inequalities. In particular, we discuss error bounds and Newton-type methods for such systems. An exhaustive comparison of various regularity conditions which arise in this context is given. We obtain a new error bound under an assumption which we show to be strictly weaker than assumptions previously used for KKT systems, such as quasi-regularity or semistability (equivalently, the R_0-property). Error bounds are useful, among other things, for identifying active constraints and developing efficient local algorithms. We propose a family of local Newton-type algorithms. This family contains some known active-set Newton methods, as well as some new methods. Regularity conditions required for local superlinear convergence compare favorably with convergence conditions of nonsmooth Newton methods and sequential quadratic programming methods.

Key words. KKT system, regularity, error bound, active constraints, Newton method.
AMS subject classifications. 90C30, 65K05.

---

* Research of the first author is supported by the Russian Foundation for Basic Research Grants 99-01-00472 and 01-01-00810. The second author is supported in part by CNPq Grant 300734/95-6, by PRONEX–Optimization, and by FAPERJ.

† Computing Center of the Russian Academy of Sciences, Vavilova Str. 40, Moscow, 117967, Russia. Email: `izmaf@ccas.ru`

‡ Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: `solodov@impa.br`

1 Introduction

In this paper we consider the *Karush–Kuhn–Tucker* (KKT) *system* of finding $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ such that

$$\begin{array}{l} \Psi(x, \mu) = 0, \\ \mu \geq 0, \; G(x) \geq 0, \; \langle \mu, G(x) \rangle = 0, \end{array} \tag{1.1}$$

where

$$\Psi : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n, \quad \Psi(x, \mu) = \Phi(x) - (G'(x))^{\mathrm{T}} \mu, \tag{1.2}$$

and $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ are given sufficiently smooth mappings. As is well-known [11], under appropriate constraint qualifications, conditions (1.1) represent the primal-dual characterization of solutions of the following variational inequality: Find $x \in D$ such that

$$\begin{array}{l} \langle \Phi(x), \xi - x \rangle \geq 0 \quad \forall \xi \in D, \\ D = \{x \in \mathbf{R}^n \mid G(x) \geq 0\}. \end{array}$$

If further for some smooth function $f : \mathbf{R}^n \to \mathbf{R}$ it holds that

$$\Phi(x) = f'(x), \quad x \in \mathbf{R}^n, \tag{1.3}$$

then (1.1) is the KKT system for the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & G(x) \geq 0. \end{array} \tag{1.4}$$

We note that all developments given below extend to the case when equality constraints are present in a straightforward manner.

In this paper, our objective is two-fold. First, we study conditions which guarantee the existence of a certain *error bound*, i.e., an upper estimate for the distance from a given point to the solution set of (1.1) in terms of some computable quantity. We obtain a new local error bound under a regularity condition which is weaker than conditions previously used in the literature in this setting (see Section 2, and in particular Figure 1). Error bounds are important by themselves, as they have a wide variety of applications [22]. Here, we use our error bound for the task of identifying active constraints, in the spirit of [5]. Having in hand this identification, we propose a new class of Newton-type algorithms with attractive local properties (see Sections 4, 5 and Figure 2). In particular, regularity conditions required for local superlinear convergence of our algorithm are weaker than conditions for other local algorithms whose iteration consists of solving one system of linear equations (e.g., standard active-set Newton methods or nonsmooth Newton methods). Furthermore, in the optimization case our method converges under the same assumptions as the sequential quadratic programming algorithm (uniqueness of the multiplier and second-order sufficiency), while its iterations are considerably simpler.

A few words about our notation. Given a finite set I, $|I|$ stands for its cardinality. By $\mathbf{R}(m, n)$ we denote the space of $m \times n$ matrices with real entries. By E we shall denote the identity matrix whose dimension would be always clear from the context. For $y \in \mathbf{R}^m$ and an index set $I \subset \{1, \ldots, m\}$, y_I stands for the vector with components y_i, $i \in I$. For a linear operator Λ, $\operatorname{im}\Lambda$ is its range (image space), and $\ker\Lambda$ is its kernel (null space). For a directionally differentiable mapping $F : \mathbf{R}^n \to \mathbf{R}^m$, by $F'(x;\, d)$ we denote the usual directional derivative of F at $x \in \mathbf{R}^n$ in the direction $d \in \mathbf{R}^n$.

2 Error Bounds for KKT Systems

We start with some known error bounds and regularity conditions. To complete the picture, we also discuss their relationships with classical constraint qualifications and second-order conditions, summarized in Figure 1. After that, we exhibit our new error bound and show that it holds under weaker assumptions.

Given a solution $(\bar{x},\, \bar{\mu}) \in \mathbf{R}^n \times \mathbf{R}^m$ of (1.1), define the index sets

$$\begin{array}{l} I = I(\bar{x}) = \{i = 1,\, \ldots,\, m \mid G_i(\bar{x}) = 0\}, \\ N = N(\bar{x}) = \{1,\, \ldots,\, m\} \setminus I. \end{array}$$

Define further the partition $(I_0,\, I_+)$ of I, where

$$\begin{array}{l} I_0 = I_0(\bar{x},\, \bar{\mu}) = \{i \in I \mid \bar{\mu}_i = 0\}, \\ I_+ = I_+(\bar{x},\, \bar{\mu}) = \{i \in I \mid \bar{\mu}_i > 0\} = I \setminus I_0. \end{array}$$

In what follows, we omit the dependency of those index sets on $\bar{x}$ and $\bar{\mu}$, as it is clear from the context. The index set I is referred to as the set of active constraints, I_+ is the set of strongly active constraints, and I_0 is the degenerate set. The *strict complementarity condition* at $(\bar{x},\, \bar{\mu})$ consists of saying that $I_0 = \emptyset$. We emphasize that we do *not* assume this condition anywhere in this paper.

As is well known, solutions of (1.1) coincide with zeroes of the following mapping based on the *natural residual* reformulation of complementarity conditions in (1.1):

$$R : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m, \quad R(x,\, \mu) = \begin{pmatrix} \Psi(x,\, \mu) \\ \min\{\mu_1,\, G_1(x)\} \\ \cdots \\ \min\{\mu_m,\, G_m(x)\} \end{pmatrix}. \tag{2.1}$$

Consider the error bound

$$\|(x - \bar{x},\, \mu - \bar{\mu})\| \le M\|R(x,\, \mu)\| \quad \forall\, (x,\, \mu) \in U, \tag{2.2}$$

where $M > 0$ is some constant and U is some neighborhood of $(\bar{x},\, \bar{\mu})$. It is known [23] that, in this context, the bound (2.2) is equivalent to the R_0*-property* of the mapping R at $(\bar{x},\, \bar{\mu})$, that is

$$R'((\bar{x},\, \bar{\mu});\, (x,\, \mu)) = 0 \Longleftrightarrow (x,\, \mu) = 0,$$

which is further equivalent to *semistability* of $(\bar{x},\, \bar{\mu})$, as defined in [2]. We note that the equivalence between the R_0-property and semistability is easy to see by defining

$$\begin{array}{rcl} Q & := & \{(x,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \mid R'((\bar{x},\, \bar{\mu});\, (x,\, \mu)) = 0\} \\ & = & \left\{ (x,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \;\middle|\; \begin{array}{l} \frac{\partial \Psi}{\partial x}(\bar{x},\, \bar{\mu})x - (G'(\bar{x}))^{\mathrm{T}}\mu = 0, \\ \min\{\mu_i,\, \langle G_i'(\bar{x}),\, x\rangle\} = 0,\, i \in I_0, \\ G'_{I_+}(\bar{x})x = 0, \\ \mu_N = 0 \end{array} \right\}, \end{array} \tag{2.3}$$

so that the R_0-property means that $Q = \{0\}$, and comparing this with [2, Proposition 5.1].

For any locally Lipschitzian mapping, the R_0-property is implied by *BD-regularity.* In its turn, BD-regularity of the mapping R at $(\bar{x}, \bar{\mu})$ (i.e., nonsingularity of every $\Lambda \in \partial_B R(\bar{x}, \bar{\mu})$, where $\partial_B R$ stands for the B-subdifferential of R) is implied by the so-called *quasi-regularity* property introduced in [5]. The latter consists of saying that the matrix

$$\begin{pmatrix} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu}) & (G'_J(\bar{x}))^{\mathrm{T}} & (G'_{I_+}(\bar{x}))^{\mathrm{T}} \\ G'_J(\bar{x}) & 0 & 0 \\ G'_{I_+}(\bar{x}) & 0 & 0 \end{pmatrix} \tag{2.4}$$

is nonsingular for every index set $J \subset I_0$. In fact, although this was not stated in [5], it can be seen that quasi-regularity is actually equivalent to BD-regularity of R at $(\bar{x}, \bar{\mu})$ (we shall not prove this formally here, but this fact would actually be quite clear from the discussion in Section 5). In its turn, quasi-regularity is a weaker property than Robinson's *strong regularity* [28]. Strong regularity means here that the determinants of matrices defined in (2.4) have the same nonzero sign for all $J \subset I_0$ [18].

Before proceeding, we state some constraint qualifications that will be used in the sequel:

- The *linear independence constraint qualification* (LICQ): $\operatorname{rank} G'_I(\bar{x}) = |I|$.
- The *strict Mangasarian–Fromovitz constraint qualification* (SMFCQ): $\operatorname{rank} G'_{I_+}(\bar{x}) = |I_+|$ and there exists $h \in \mathbf{R}^n$ such that $G'_{I_0}(\bar{x})h > 0$, $G'_{I_+}(\bar{x})h = 0$.
- The *weak linear independence constraint qualification* (WLICQ): $\operatorname{rank} G'_{I_+}(\bar{x}) = |I_+|$.

Note that SMFCQ is equivalent to uniqueness of the multiplier $\bar{\mu}$ associated with $\bar{x}$ [21]. We shall use the following (equivalent) dual form of SMFCQ: for $\mu \in \mathbf{R}^m$ it holds that

$$\left.\begin{array}{r} (G'_I(\bar{x}))^{\mathrm{T}} \mu_I = 0 \\ \mu_{I_0} \geq 0 \\ \mu_{I_N} = 0 \end{array}\right\} \quad \Longleftrightarrow \quad \mu = 0. \tag{2.5}$$

Also, we shall employ certain second-order conditions of the form

$$\left\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, \, x \right\rangle \neq 0 \quad \forall\, x \in K \setminus \{0\},$$

with different choices of the cone $K \subset \mathbf{R}^n$:

- The *strong second-order sufficiency condition* (SSOSC): $K = \{x \in \mathbf{R}^n \mid G'_{I_+}(\bar{x})x = 0\}$.
- The *second-order sufficiency condition* (SOSC): $K = \{x \in \mathbf{R}^n \mid G'_{I_0}(\bar{x})x \geq 0, \; G'_{I_+}(\bar{x})x = 0\}$.
- The *second-order condition* (SOC): $K = \{x \in \mathbf{R}^n \mid G'_I(\bar{x})x = 0\}$.

Note that all these second-order conditions subsume that $\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, \, x \rangle$ has the same sign for all x in the corresponding K.

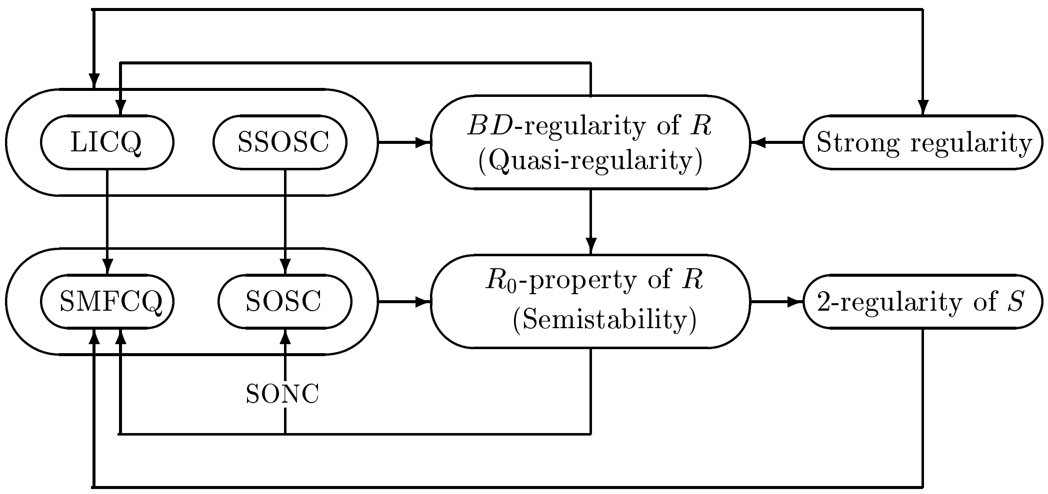

Figure 1: BD-regularity and R_0-property of R, and 2-regularity of S

Finally, the *second-order necessity condition* (SONC) consists of saying that

$$\left\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, x \right\rangle \geq 0 \quad \forall\, x \in K,$$

with K defined as in SOSC.

The words "sufficiency" and "necessity" are used here by analogy with the optimization case, where those conditions are part of the sufficient/necessary conditions for $\bar{x}$ to be a local minimizer. For a general KKT system, this terminology should not be taken literally.

Relations between the R_0-property, quasi-regularity, and classical constraint qualifications and second-order conditions are displayed in Figure 1. As established in [5], quasi-regularity implies LICQ, and is implied by LICQ combined with SSOSC, but SSOSC is not implied by quasi-regularity. In the optimization case, by the results in [28, 18] (see also [3]) it further holds that strong regularity is equivalent to LICQ combined with SSOSC. We next establish relationships with the R_0-property (semistability). The following result can be essentially deduced from [2, Proposition 6.2], but it was stated for the optimization case only. We extend it to the general KKT system, and include a proof for completeness.

Proposition 2.1 *For the mapping R defined by* (2.1)*, the R_0-property of the mapping R at $(\bar{x}, \bar{\mu})$ is implied by SMFCQ combined with SOSC. The R_0-property implies SMFCQ. Under SONC, if $\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})$ is symmetric, the R_0-property also implies SOSC.*

Proof. We first prove that SOSC and SMFCQ imply the R_0-property. Take an arbitrary

$(x, \mu) \in Q$. By (2.3), there exists a partition (J_1, J_2) of I_0 such that

$$\begin{array}{l} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x - (G'(\bar{x}))^{\mathrm{T}}\mu = 0, \\ \mu_{J_1} \geq 0, \; G'_{J_1}(\bar{x})x = 0, \\ \mu_{J_2} = 0, \; G'_{J_2}(\bar{x})x \geq 0, \\ G'_{I_+}(\bar{x})x = 0, \\ \mu_N = 0. \end{array} \tag{2.6}$$

Clearly, $x \in K$ defined in SOSC. Furthermore, $\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, x\rangle = \langle \mu, G'(\bar{x})x\rangle = 0$. Hence, under SOSC, it must be the case that $x = 0$. Under SMFCQ, (2.6) and (2.5) now imply that $\mu = 0$. Hence, the R_0-property holds.

On the other hand, suppose that SMFCQ is violated, i.e., there exists $\mu \in \mathbf{R}^m \setminus \{0\}$ satisfying relations in the left-hand side of (2.5). Then $(0, \mu)$ satisfies (2.6) with $J_1 = I_0$, $J_2 = \emptyset$, i.e., $(0, \mu) \in Q \setminus \{0\}$, which means that the R_0-property is violated. It follows that the R_0-property implies SMFCQ.

Finally, let SOSC be violated, i.e., there exists $x \in K \setminus \{0\}$ such that $\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, x\rangle = 0$. Then, under SONC, x is a solution of the linearly-constrained optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})\xi, \xi\rangle \\ \text{subject to} & \xi \in K. \end{array}$$

In particular, x satisfies the KKT optimality system for this problem. If $\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})$ is symmetric, the latter means that there exists some $\mu_I \in \mathbf{R}^m$ such that

$$\begin{array}{l} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x - (G'_{I_0}(\bar{x}))^{\mathrm{T}}\mu_{I_0} - (G'_{I_+}(\bar{x}))^{\mathrm{T}}\mu_{I_+} = 0, \\ \mu_{I_0} \geq 0, \; G'_{I_0}(\bar{x})x \geq 0, \; \langle \mu_{I_0}, G'_{I_0}(\bar{x})x\rangle = 0, \\ G'_{I_+}(\bar{x})x = 0. \end{array}$$

Then (x, μ) with given μ_I and $\mu_N = 0$ satisfies (2.6) with $J_1 = \{i \in I_0 \mid \langle G'_i(\bar{x}), x\rangle = 0\}$, $J_2 = I_0 \setminus J_1$. We therefore have that $(x, \mu) \in Q \setminus \{0\}$, and the R_0-property is violated. Hence, under the given assumptions, the R_0-property implies SOSC. This completes the proof. ∎

Note that $\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})$ is automatically symmetric in the optimization case (1.3), and SONC also holds in that case if further $\bar{x}$ is a local minimizer of (1.4).

Proposition 2.2 *For the mapping R defined in* (2.1)*, the R_0-property at $(\bar{x}, \bar{\mu})$ is implied by quasi-regularity, but not vice versa.*

Proof. Indeed, for an arbitrary $(x, \mu) \in Q$ take the index set $J = J_1$ with J_1 as in (2.6). Then $(x, \mu_{J \cup I_+})$ evidently belongs to the null space of the matrix defined in (2.4). This shows that quasi-regularity implies the R_0-property. The inverse implication does not hold, as illustrated by the following example. ∎

Example 2.1 Let $n = m = 2$, Φ be defined by (1.3) with $f(x) = (x_1 + x_2)^2/2$, $G(x) = (x_1,\, x_2)$, $x \in \mathbf{R}^2$. Then $\bar{x} = 0$ is a solution of (1.4) (actually, it is an unconstrained global minimizer of f), and the unique associated multiplier is $\bar{\mu} = 0$. We have that $I_0 = \{1,\, 2\}$, $I_+ = N = \emptyset$. Quasi-regularity is violated (for $J = \emptyset$). At the same time, LICQ (and hence, SMFCQ) and SOSC (but not SSOSC) are satisfied in this example. The R_0-property follows now from Proposition 2.1.

We next exhibit a new error bound based on the theory of *2-regularity* developed in [13, 14]. We further show that conditions which are needed for it to hold are strictly weaker than conditions discussed above.

Consider the following mapping given by a *smooth* reformulation of complementarity conditions:

$$S : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m, \quad S(x,\, \mu) = \begin{pmatrix} \Psi(x,\, \mu) \\ 2\mu_1 G_1(x) - (\min\{0,\, \mu_1 + G_1(x)\})^2 \\ \cdots \\ 2\mu_m G_m(x) - (\min\{0,\, \mu_m + G_m(x)\})^2 \end{pmatrix}. \tag{2.7}$$

As is well-known [17, 29] and easy to see, zeroes of S also coincide with solutions of the KKT system (1.1). But unlike the natural residual $R(\cdot)$, the mapping $S(\cdot)$ is smooth. On the other hand, if strict complementarity does not hold then S is necessarily irregular at $(\bar{x},\, \bar{\mu})$ in the sense that the matrix $S'(\bar{x},\, \bar{\mu})$ is singular. Nevertheless, the singularity is "structured", which allows a certain form of second-order analysis [13, 14].

Let P be the orthogonal projector onto $(\operatorname{im} S'(\bar{x},\, \bar{\mu}))^\perp$. (Actually, one could take any complementary subspace Y of $\operatorname{im} S'(\bar{x},\, \bar{\mu})$, and define P as the projector onto Y parallel to $\operatorname{im} S'(\bar{x},\, \bar{\mu})$.) Then under appropriate smoothness assumptions on Φ and G, S is continuously differentiable at $(\bar{x},\, \bar{\mu})$, PS' is Lipschitz-continuous in some neighbourhood of $(\bar{x},\, \bar{\mu})$ and directionally differentiable at this point with respect to every direction. Since the mapping S is acting from the space $\mathbf{R}^n \times \mathbf{R}^m$ into itself, the notion of 2-regularity [13, 14] in that special case reduces to the following.

Definition 2.1 The mapping S defined in (2.7) is 2-*regular* at $(\bar{x},\, \bar{\mu})$ if

$$T := \{(x,\, \mu) \in \ker S'(\bar{x},\, \bar{\mu}) \mid (PS')'((\bar{x},\, \bar{\mu});\, (x,\, \mu))(x,\, \mu) = 0\} = \{0\}. \tag{2.8}$$

The following is a direct consequence of [14, Theorem 4].

Theorem 2.1 *Let mappings $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ be sufficiently smooth near a point $\bar{x} \in \mathbf{R}^n$, and let $(\bar{x},\, \bar{\mu})$ with some $\bar{\mu} \in \mathbf{R}^m$ be a solution of the KKT system* (1.1) *with Ψ given by* (1.2). *Suppose further that the mapping S defined in* (2.7) *is 2-regular at $(\bar{x},\, \bar{\mu})$.*

Then there exist a neighborhood U of $(\bar{x},\, \bar{\mu})$ and a number $M > 0$ such that

$$\|(x - \bar{x},\, \mu - \bar{\mu})\| \le M(\|(E - P)S(x,\, \mu)\| + \|PS(x,\, \mu)\|^{1/2}) \quad \forall\, (x,\, \mu) \in U. \tag{2.9}$$

In particular (possibly with different U and M),

$$\|(x - \bar{x},\, \mu - \bar{\mu})\| \le M\|S(x,\, \mu)\|^{1/2} \quad \forall\, (x,\, \mu) \in U. \tag{2.10}$$

Note that the error bound (2.10) is implied by error bound (2.2) (this follows from a comparison of growth rates in [29]). However, (2.10) can hold when (2.2) does not, see Example 2.2 below. In addition, the more accurate estimate (2.9) does not follow from (2.2). We next establish that 2-regularity is a weaker requirement than the R_0-property.

Proposition 2.3 *Let the mappings R and S be defined by* (2.1) *and* (2.7)*, respectively. The R_0-property of R at $(\bar{x}, \bar{\mu})$ implies 2-regularity of S at $(\bar{x}, \bar{\mu})$, but not vice versa.*

Proof. By direct computations, we have that

$$S'(\bar{x}, \bar{\mu}) = \begin{pmatrix} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu}) & -(G'(\bar{x}))^{\mathrm{T}} \\ A & B \end{pmatrix},$$

where $A \in \mathbf{R}(m, n)$ and $B \in \mathbf{R}(m, m)$ are given by

$$A_i = \begin{cases} 0, & i \in I_0 \cup N, \\ \bar{\mu}_i G_i'(\bar{x}), & i \in I_+, \end{cases} \qquad B_i = \begin{cases} 0, & i \in I, \\ G_i(\bar{x}) e^i, & i \in N, \end{cases}$$

with $e^1, \ldots, e^m$ being the canonical basis in $\mathbf{R}^m$. Then P, the orthogonal projector onto $(\operatorname{im} S'(\bar{x}, \bar{\mu}))^{\perp}$ in $\mathbf{R}^m$, satisfies

$$(Py)_i = y_i \quad \forall\, i \in I_0,\, \forall\, y \in \mathbf{R}^m.$$

Since $PS' = (PS)'$, from (2.8) it now easily follows that

$$T \subset \{(x, \mu) \in \ker S'(\bar{x}, \bar{\mu}) \mid (S_i')'((\bar{x}, \bar{\mu});\, (x, \mu))(x, \mu) = 0,\ i \in I_0\}. \tag{2.11}$$

Clearly,

$$\ker S'(\bar{x}, \bar{\mu}) = \left\{ (x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \;\middle|\; \begin{array}{l} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x - (G'(\bar{x}))^{\mathrm{T}}\mu = 0, \\ G_{I_+}'(\bar{x})x = 0, \\ \mu_N = 0 \end{array} \right\}. \tag{2.12}$$

Furthermore, for $i \in I_0$, we have

$$\begin{aligned} (S_i')'((\bar{x}, \bar{\mu});\, (x, \mu)) \;&=\; 2(\mu_i - \min\{0,\, \langle G_i'(\bar{x}),\, x\rangle + \mu_i\})(G_i'(\bar{x}),\, 0) \\ &\quad + 2(\langle G_i'(\bar{x}),\, x\rangle - \min\{0,\, \langle G_i'(\bar{x}),\, x\rangle + \mu_i\})(0,\, e^i), \end{aligned}$$

and

$$(S_i')'((\bar{x}, \bar{\mu});\, (x, \mu))(x, \mu) = 2(2\mu_i \langle G_i'(\bar{x}),\, x\rangle - (\min\{0,\, \langle G_i'(\bar{x}),\, x\rangle + \mu_i\})^2).$$

Observe that the right-hand side of the latter equality is zero if, and only if, the following complementarity conditions are satisfied:

$$\mu_i \geq 0,\ \langle G_i'(\bar{x}),\, x\rangle \geq 0,\ \mu_i \langle G_i'(\bar{x}),\, x\rangle = 0, \quad i \in I_0.$$

Combining this fact with (2.12) and comparing (2.11) with (2.3), it is now evident that $T \subset Q$.

Recall that $Q = \{0\}$ is equivalent to the R_0-property. Hence, the R_0-property is a sufficient condition for 2-regularity of S at $(\bar{x}, \bar{\mu})$.

The following example demonstrates that 2-regularity can hold without the R_0-property. ■

Example 2.2 Let $n = m = 2$, Φ be defined by (1.3) with $f(x) = x_1^2/2 + x_2^3/3$, $G(x) = (x_1 - x_2^2/2,\, x_1 + x_2^2/2)$, $x \in \mathbf{R}^2$. Then $(\bar{x}, \bar{\mu})$ with $\bar{x} = 0$ and $\bar{\mu} = 0$ is a solution of (1.1), $I_0 = \{1, 2\}$, $I_+ = N = \emptyset$. This solution satisfies SMFCQ and SONC, but not SOSC. Because under SONC the R_0-property implies SOSC (recall Proposition 2.1), it follows that the R_0-property does not hold here. At the same time,

$$T = \left\{ (x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \;\middle|\; \begin{array}{l} x_1 - \mu_1 - \mu_2 = 0, \\ x_2^2 + (\mu_1 - \mu_2)x_2 = 0, \\ \min\{\mu_1, x_1\} = 0, \\ \min\{\mu_2, x_1\} = 0 \end{array} \right\} = \{0\},$$

and 2-regularity holds, together with error bounds (2.9), (2.10).

Observe that R indeed fails to provide the error bound (2.2): taking the sequences $\{(x^k, \mu^k)\} \subset \mathbf{R}^2 \times \mathbf{R}$ with $x_1^k = 0$, $x_2^k = 1/k$ and $\mu^k = 0$, we have that $\|(x^k, \mu^k) - (\bar{x}, \bar{\mu})\| = 1/k$ but $\|R(x^k, \mu^k)\| = \sqrt{5}/(2k^2)\ \forall\, k$.

In the given example $\bar{x}$ is a KKT point, but not a local solution of the optimization problem (1.4). However, nothing would change if we choose $f(x) = x_1^2/2 + |x_2|^3/3$, $x \in \mathbf{R}^2$, leaving the other data as is. With this choice, $\bar{x}$ would become a solution of the optimization problem, the R_0-property does not hold, but 2-regularity does.

As an aside, note that if

$$\operatorname{rank} \begin{pmatrix} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu}) & (G'_{I_0}(\bar{x}))^{\mathrm{T}} & (G'_{I_+}(\bar{x}))^{\mathrm{T}} \\ G'_{I_+}(\bar{x}) & 0 & 0 \end{pmatrix} = n + |I_+| \tag{2.13}$$

then the projector P is easily explicitly computable, and it can be seen that (under this condition which, in particular, is implied by quasi-regularity), 2-regularity and the R_0-property are equivalent, as the cone T in (2.8) coincides with Q in (2.3).

For an arbitrary $\mu \in \mathbf{R}^m$ satisfying relations in the left-hand side of (2.5), it can be easily seen that $(0, \mu) \in T$. Hence, 2-regularity of S at $(\bar{x}, \bar{\mu})$ subsumes SMFCQ. Generally speaking, conditions needed for all error bounds discussed in this section imply that $(\bar{x}, \bar{\mu})$ is an isolated solution of (1.1), and in particular, multiplier $\bar{\mu}$ associated with $\bar{x}$ is unique (i.e., SMFCQ holds).

3 Identification of Active Constraints and a Simple Active-Set Newton Method

An important application of error bounds for KKT systems is identifying active constraints, i.e., the index set I. This identification is useful for many purposes, and it will be also employed for developing our class of Newton-type algorithms in Section 4. We next briefly review an identification technique, which is due to [5]. We also discuss a simple active-set Newton method, which is not original by itself, but it would be useful for comparison purposes with our algorithm.

Let $(\bar{x}, \bar{\mu})$ be an isolated solution of (1.1). We note that for identification technique of [5] the multiplier does not have to be unique. On the other hand, SMFCQ is subsumed by conditions that will be needed later on, and so we shall assume the uniqueness from the beginning. Define

$$I(x, \mu) := \{i = 1, \ldots, m \mid \rho(\|S(x, \mu)\|) \geq G_i(x)\}, \quad x \in \mathbf{R}^n, \mu \in \mathbf{R}^m, \tag{3.1}$$

where

$$\rho : \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = \begin{cases} -1/\log t, & \text{if } t > 0, \\ 0, & \text{if } t = 0. \end{cases}$$

It can be verified that if the error bound (2.10) holds, then $I(x, \mu)$ coincides with I for *all* (x, μ) sufficiently close to $(\bar{x}, \bar{\mu})$. Of course, the mapping R can be used instead of S, provided (2.2) is satisfied. However, as shown above, this would require stronger assumptions. Finally, one could also use

$$\rho : \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = t^\theta,$$

with $\theta \in (0, 1/2)$ for S, and $\theta \in (0, 1)$ for R.

An attractive feature of this approach is that it can be incorporated within any algorithm. Furthermore, the only requirement here is a local error bound. Assumptions which guarantee the bound (2.2) (which are stronger than for (2.10)) are already weaker than assumptions needed in other methods of identification, see [5] for a discussion.

Once I is identified, probably the most direct way to employ this information is to replace (1.1) by the pure system of equations

$$\begin{aligned} &\Phi(x) - (G_I'(x))^{\mathrm{T}} \mu_I = 0, \\ &G_I(x) = 0, \end{aligned} \tag{3.2}$$

or equivalently,

$$F(x, \mu_I) = 0$$

with

$$F : \mathbf{R}^n \times \mathbf{R}^{|I|} \to \mathbf{R}^n \times \mathbf{R}^{|I|}, \quad F(x, \mu_I) = \begin{pmatrix} \Phi(x) - (G_I'(x))^{\mathrm{T}} \mu_I \\ G_I(x) \end{pmatrix}. \tag{3.3}$$

Indeed, $(\bar{x}, \bar{\mu}_I)$ is a solution of (3.2), and this solution can now be found by the Newton-type methods for smooth nonlinear equations, provided

$$F'(\bar{x}, \bar{\mu}_I) = \begin{pmatrix} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu}) & -(G_I'(\bar{x}))^{\mathrm{T}} \\ G_I'(\bar{x}) & 0 \end{pmatrix} \tag{3.4}$$

is nonsingular. We next establish the relationships between the latter condition and other regularity concepts used in this paper. These will further be compared with assumptions needed for the new class of methods introduced in Section 4, see Figure 2.

Proposition 3.1 *For the mapping F defined in (3.3), nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ is implied by LICQ combined with SOC. Nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ implies LICQ.*

Proof. Take an arbitrary $(x, \mu_I) \in \ker F'(\bar{x}, \bar{\mu}_I)$. According to (3.4), this means that

$$\begin{array}{l} \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x - (G_I'(\bar{x}))^{\mathrm{T}}\mu_I = 0, \\ G_I'(\bar{x})x = 0. \end{array} \tag{3.5}$$

In particular, $x \in K$ defined in SOC. Furthermore, $\langle \frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})x, x\rangle = \langle \mu_I, G_I'(\bar{x})x\rangle = 0$. Under SOC, it must be the case that $x = 0$. Under LICQ, it then follows from (3.5) that $\mu_I = 0$. Hence, $F'(\bar{x}, \bar{\mu}_I)$ is nonsingular.

The fact that nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ subsumes LICQ is evident from (3.4). ∎

Obviously, quasi-regularity (equivalently, BD-regularity of R) also implies nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$, but not vice versa. The latter is illustrated by Example 2.1. Indeed, quasi-regularity does not hold in this example, but LICQ and SOC do hold, and hence $F'(\bar{x}, \bar{\mu}_I)$ is nonsingular, by Proposition 3.1.

The R_0-property of R at $(\bar{x}, \bar{\mu})$ and the nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ are not directly related: neither of them is implied by the other. In particular, the R_0-property can hold without LICQ which is, by Proposition 3.1, a necessary condition for nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$. Those facts are illustrated by the following two examples.

Example 3.1 Let $n = 2$, $m = 3$, Φ be defined by (1.3) with $f(x) = x_1 + (x_1^2 + x_2^2)/2$, $G(x) = (x_1, x_2, x_1 + x_2)$, $x \in \mathbf{R}^2$. Then $\bar{x} = 0$ is a solution of (1.4) with the unique associated multiplier $\bar{\mu} = (1, 0, 0)$, and $I_0 = \{2, 3\}$, $I_+ = \{1\}$, $N = \emptyset$. LICQ does not hold. Hence, $F'(\bar{x}, \bar{\mu}_I)$ is singular. At the same time, SMFCQ and SSOSC (hence, SOSC) are satisfied in this example. The R_0-property follows now from Proposition 2.1.

Example 3.2 Let $n = m = 1$, Φ be defined by (1.3) with $f(x) = x^4/4$, $G(x) = x$, $x \in \mathbf{R}$. Then $\bar{x} = 0$ is a solution of (1.4) with the unique associated multiplier $\bar{\mu} = 0$, and $I_0 = \{1\}$, $I_+ = N = \emptyset$. It is easy to verify by direct computations that $F'(\bar{x}, \bar{\mu}_I)$ is nonsingular, but the R_0-property does not hold.

However, both the R_0-property and the nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ are implied by quasi-regularity (see Proposition 2.2 and the discussion above). Moreover, they can be both satisfied simultaneously without quasi-regularity, as illustrated by Example 2.1.

4 A New Class of Newton-Type Methods for Solving KKT Systems

A technique similar to the one used in Section 3 for identifying active constraints I can also be used to identify the degenerate set I_0 (and hence, the set of strongly active constraints

$I_+ = I \setminus I_0$) [5]. In particular, the index set

$$\begin{array}{rcl} I_0(x,\, \mu) & = & \{i \in I(x,\, \mu) \mid \rho(\|S(x,\, \mu)\|) \geq \mu_i\} \\ & = & \{i = 1,\, \ldots,\, m \mid \rho(\|S(x,\, \mu)\|) \geq \max\{\mu_i,\, G_i(x)\}\}, \\ & & x \in \mathbf{R}^n,\, \mu \in \mathbf{R}^m, \end{array} \tag{4.1}$$

coincides with I_0 for all $(x,\, \mu)$ sufficiently close to $(\bar{x},\, \bar{\mu})$, provided the error bound (2.10) holds. Below we demonstrate what advantages can be extracted from this information when $I_0 \neq \emptyset$. Note that when strict complementarity holds ($I_0 = \emptyset$), arguably all Newton-type methods for solving the KKT systems (e.g., active-set Newton methods, the nonsmooth Newton methods, sequential quadratic programming, etc.) should reduce to the same pure Newton method for solving smooth nonlinear equations. In this sense, the strict complementarity case is not interesting. As usual in this paper, below we do not assume strict complementarity.

Once I_0 is identified, the most obvious way to use this information is to reduce the number of variables in (3.2) by setting $\mu_{I_0} = 0$, and removing from (3.2) equations corresponding to $i \in I_0$:

$$\begin{array}{l} \Phi(x) - (G'_{I_+}(x))^{\mathrm{T}} \mu_{I_+} = 0, \\ G_{I_+}(x) = 0. \end{array} \tag{4.2}$$

Our proposal is to use the information about I_0 in a more general way, by introducing a system enlarged by a certain parametrization. The key advantage is that this would lead to weaker regularity conditions needed for local superlinear/quadratic convergence. Additionally, this approach presents a convenient unification of various active-set methods, among which are the ones based on (3.2) and (4.2). In some sense, even the nonsmooth Newton methods and sequential quadratic programming can be related to the framework proposed below, see Section 5.

Consider the following parametric system of equations:

$$\begin{array}{l} \Phi(x) - (G'_I(x))^{\mathrm{T}} \mu_I = 0, \\ AG_{I_0}(x) + B\mu_{I_0} = 0, \\ G_{I_+}(x) = 0, \end{array} \tag{4.3}$$

where $A,\, B \in \mathbf{R}(|I_0|,\, |I_0|)$ are parameter matrices. Clearly, $(\bar{x},\, \bar{\mu}_I)$ is a solution of (4.3) with any choice of A and B. Formally, (4.2) corresponds to the specific choice $A = 0$, $B = E$ in (4.3), while (3.2) corresponds to $A = E$, $B = 0$.

For a pair of matrices $A,\, B \in \mathbf{R}(|I_0|,\, |I_0|)$, define the mapping

$$F_{A,\, B} : \mathbf{R}^n \times \mathbf{R}^{|I|} \to \mathbf{R}^n \times \mathbf{R}^{|I|}, \quad F_{A,\, B}(x,\, \mu_I) = \begin{pmatrix} \Phi(x) - (G'_I(x))^{\mathrm{T}} \mu_I \\ AG_{I_0}(x) + B\mu_{I_0} \\ G_{I_+}(x) \end{pmatrix}, \tag{4.4}$$

so that (4.3) becomes

$$F_{A,\, B}(x,\, \mu_I) = 0.$$

We can now attempt to solve this system of equations by standard Newton-type methods, provided it satisfies the usual conditions. In particular, the matrix

$$F'_{A,\,B}(\bar{x},\,\bar{\mu}_I)=\left(\begin{array}{ccc}\frac{\partial\Psi}{\partial x}(\bar{x},\,\bar{\mu}) & -(G'_{I_0}(\bar{x}))^{\mathrm{T}} & -(G'_{I_+}(\bar{x}))^{\mathrm{T}}\\ AG'_{I_0}(\bar{x}) & B & 0\\ G'_{I_+}(\bar{x}) & 0 & 0\end{array}\right), \tag{4.5}$$

should be nonsingular.

We could immediately consider some specific choices of A and B, and derive conditions which are needed for nonsingularity of $F'_{A,\,B}(\bar{x},\,\bar{\mu}_I)$ for those fixed choices. But, as we shall see, there is an advantage in treating A and B as parameters. Furthermore, conditions which are required for certain specific choices of A and B will be clear from the analysis below.

We start with the following definition.

Definition 4.1 A solution $(\bar{x},\,\bar{\mu})$ of the KKT system (1.1) is referred to as *weakly regular* if there exists a pair of matrices $A,\,B\in\mathbf{R}(|I_0|,\,|I_0|)$ such that for $F_{A,\,B}$ defined in (4.4), the matrix $F'_{A,\,B}(\bar{x},\,\bar{\mu}_I)$ is nonsingular.

Proposition 4.1 *Suppose that the solution* $(\bar{x},\,\bar{\mu})$ *of the KKT system* (1.1) *is weakly regular.*

Then the set of pairs $A,\,B\in\mathbf{R}(|I_0|,\,|I_0|)$ *such that* $F'_{A,\,B}(\bar{x},\,\bar{\mu}_I)$ *is nonsingular is open and dense in* $\mathbf{R}(|I_0|,\,|I_0|)\times\mathbf{R}(|I_0|,\,|I_0|)$.

Proof. The determinant $\det F'_{A,\,B}(\bar{x},\,\bar{\mu}_I)$ is a polynomial with respect to the entries of $(A,\,B)\in\mathbf{R}(|I_0|,\,|I_0|)\times\mathbf{R}(|I_0|,\,|I_0|)$. By weak regularity (see Definition 4.1), this polynomial is not everywhere zero. Hence, the set where this polynomial is not zero is obviously open and dense in $\mathbf{R}(|I_0|,\,|I_0|)\times\mathbf{R}(|I_0|,\,|I_0|)$. ∎

Proposition 4.1 means that we can think of weak regularity as the condition required for solving (4.3) by conventional Newton-type methods. Indeed, since the set of "good" A and B is open and dense (its closure is the whole space), the set of "bad" choices is "thin". Therefore, A and B can be essentially arbitrary. If stronger than weak regularity (but still quite mild) conditions are assumed, then specific choices of parameters can be easily exhibited, as will be shown below.

We next study the condition of weak regularity. The relations between weak regularity, nonsingularity of $F'(\bar{x},\,\bar{\mu}_I)$, BD-regularity of R (quasi-regularity) and classical constraint qualifications and second-order conditions are summarized in Figure 2.

Lemma 4.1 *Let* $C\in\mathbf{R}(n,\,n)$, $D_1\in\mathbf{R}(m_1,\,n)$, *and* $D_2\in\mathbf{R}(m_2,\,n)$ *be such that*

$$\operatorname{rank} D_2=m_2, \tag{4.6}$$

$$\langle Cx,\,x\rangle>0\quad\forall\,x\in(\ker D_1\cap\ker D_2)\setminus\{0\}. \tag{4.7}$$

Then there exists $\bar{c}\geq 0$ *such that for all* $c>\bar{c}$ *the matrix*

$$\left(\begin{array}{ccc}C & -D_1^{\mathrm{T}} & -D_2^{\mathrm{T}}\\ cD_1 & E & 0\\ D_2 & 0 & 0\end{array}\right) \tag{4.8}$$

is nonsingular.

Proof. By (4.7) (e.g., [1, Lemma 3.2.1]), there exists $\bar{c} \geq 0$ such that for all $c > \bar{c}$ it holds that

$$\langle (C + c(D_1^{\mathrm{T}} D_1 + D_2^{\mathrm{T}} D_2))x,\, x\rangle > 0 \quad \forall\, x \in \mathbf{R}^n \setminus \{0\}.$$

In particular,

$$\langle (C + cD_1^{\mathrm{T}} D_1)x,\, x\rangle > 0 \quad \forall\, x \in \ker D_2 \setminus \{0\}. \tag{4.9}$$

Take an arbitrary $(x,\, \mu^1,\, \mu^2) \in \mathbf{R}^n \times \mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$ belonging to the kernel of the matrix defined in (4.8), i.e.,

$$\begin{aligned} &Cx - D_1^{\mathrm{T}}\mu^1 - D_2^{\mathrm{T}}\mu^2 = 0,\\ &cD_1 x + \mu^1 = 0,\\ &D_2 x = 0. \end{aligned} \tag{4.10}$$

In particular, we have that $x \in \ker D_2$ and

$$\begin{aligned} 0 &= \langle Cx,\, x\rangle - \langle \mu^1,\, D_1 x\rangle - \langle \mu^2,\, D_2 x\rangle \\ &= \langle Cx,\, x\rangle + \langle cD_1 x,\, D_1 x\rangle \\ &= \langle (C + cD_1^{\mathrm{T}} D_1)x,\, x\rangle. \end{aligned}$$

It then follows from (4.9) that $x = 0$. The second equation in (4.10) yields now that $\mu^1 = 0$, and the first equation in (4.10), combined with (4.6), imply that $\mu^2 = 0$. Hence, the matrix in (4.8) is nonsingular. ■

Proposition 4.2 *Weak regularity is implied by WLICQ combined with SOC. Weak regularity implies WLICQ.*

Proof. The first assertion follows immediately from Lemma 4.1, by setting $C = \frac{\partial \Psi}{\partial x}(\bar{x},\, \bar{\mu})$, $D_1 = G'_{I_0}(\bar{x})$, $D_2 = G'_{I_+}(\bar{x})$, and taking $A = cE$ or $A = -cE$ (with $c > 0$ large enough) and $B = E$ in (4.5). The fact that weak regularity subsumes WLICQ is obvious, by (4.5). ■

We note that it is advantageous to consider the parametrized system (4.3), as it is more likely to be nonsingular than, for example, (3.2) or (4.2). In particular, (4.3) can be nonsingular even if LICQ fails (in that case (3.2) is singular) or SSOSC fails (in which case (4.2) can be singular). Those facts can be illustrated by Examples 3.1 and 3.2, as will be discussed below. Of course, if instead of weak regularity we assume stronger conditions then this information can be used for choosing A and B. For example, if we assume that WLICQ and SOC hold, then the proof of Proposition 4.2 shows that we could take $A = cE$ or $A = -cE$ for $c > 0$ large enough, and $B = E$. Under WLICQ and SSOSC, we could take arbitrary $c \geq 0$ (in particular, $A = 0$, $B = E$); and under LICQ and SOC, $A = E$, $B = 0$. However, Proposition 4.1 shows that taking more general parameters (or better said, less special) should in fact do the job without assuming those stronger conditions.

Note that weak regularity can hold without LICQ, which is the necessary condition for nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$ in (3.4). For instance, this is the case in Example 3.1: weak regularity in this example follows from WLICQ, SOC (even SSOSC), and Proposition 4.2, while LICQ is violated. Of course, weak regularity is implied by nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$, and hence also by BD-regularity of R (or quasi-regularity).

Weak regularity can also hold when the system (4.2) is degenerate at the solution. In particular, this is the case in Example 3.2: it is easily seen that WLICQ and SOC hold, while SSOSC does not, and $F'_{A,\,B}(\bar{x}, \bar{\mu}_I)$ is singular for $A = 0$, $B = E$ corresponding to (4.2), but it is nonsingular for any $A \neq 0$.

Based on the considerations above, a local algorithm can be stated as follows.

Algorithm 4.1 *Preliminary step.* Fix (arbitrarily, e.g., randomly) two matrices $A_0, B_0 \in \mathbf{R}(m, m)$, and choose some rule for selecting square submatrices (of a given size) of these matrices.

Initialization step. Given $x^0 \in \mathbf{R}^n$ and $\mu^0 \in \mathbf{R}^m$, take $I = I(x^0, \mu^0)$ according to (3.1), and $I_0 = I_0(x^0, \mu^0)$ according to (4.1). Set $N = \{1, \ldots, m\} \setminus I$ and $I_+ = I \setminus I_0$. Select $|I_0| \times |I_0|$-submatrices A and B of matrices A_0 and B_0, respectively.

Main step. Generate the sequence $\{(x^k, \mu^k)\}$ as follows: starting from the initial point (x^0, μ_I^0), use the Newton method applied to system (4.3) with $F_{A,\,B}$ defined in (4.4) to produce $\{(x^k, \mu_I^k)\}$, setting $\mu_N^k = 0\ \forall\, k$.

From our discussion in Section 3, it follows that if the error bound (2.10) holds, and if (x^0, μ^0) is sufficiently close to $(\bar{x}, \bar{\mu})$, the sets I and I_0 will be identified correctly on the initialization step of the algorithm. Hence, according to Proposition 4.1, if $(\bar{x}, \bar{\mu})$ is a weakly regular solution of the KKT system (1.1), then it is extremely unlikely to select "wrong" A and B on the initialization step (as the set of those pairs for which $F'_{A,\,B}(\bar{x}, \bar{\mu}_I)$ is singular is "thin"). Of course, as discussed above, if instead of weak regularity we assume stronger conditions then this information can be used for choosing A and B.

We note that our approach can be further generalized/modified in several ways. For example, in the first equation in (4.3) one could use the index set I_+ instead of I. It is interesting to note that we are actually extracting *advantage* from the violation of the strict complementarity condition. When $I_0 = \emptyset$, systems of equations (3.2), (4.2) and (4.3) are all the same, and so are conditions for solving them. However, if $I_0 \neq \emptyset$ then regularity conditions for solving (4.3) are the weakest.

Theorem 4.1 *Let mappings $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ and $G : \mathbf{R}^n \to \mathbf{R}^m$ be sufficiently smooth near a point $\bar{x} \in \mathbf{R}^n$, and let $\bar{\mu} \in \mathbf{R}^m$ be such that $(\bar{x}, \bar{\mu})$ is a weakly regular solution of the KKT system* (1.1) *with Ψ given by* (1.2)*. Finally, let the error bound* (2.10) *hold for S given by* (2.7)*.*

Then for almost any (in the sense of Lebesgue measure) pair $(A_0, B_0) \in \mathbf{R}(m, m) \times \mathbf{R}(m, m)$, Algorithm 4.1 converges to $(\bar{x}, \bar{\mu})$ locally superlinearly/quadratically.

It follows that sufficient conditions for convergence of Algorithm 4.1 are weak regularity and 2-regularity. In Section 5 we show that this is weaker than conditions required for

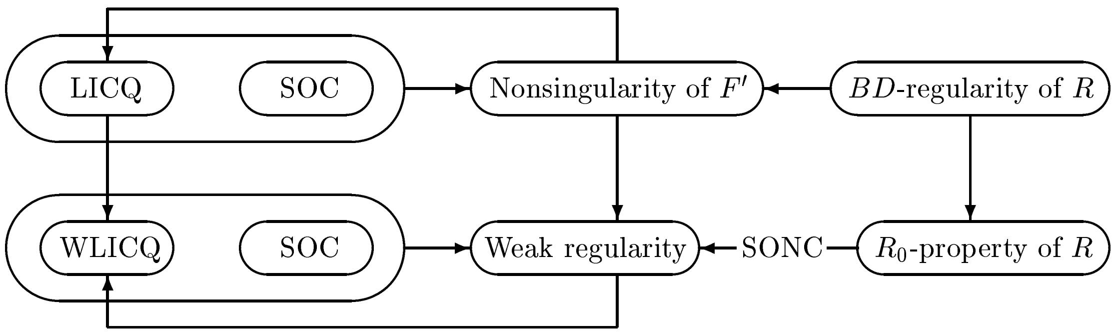

Figure 2: Nonsingularity of F', weak regularity, and BD-regularity and R_0-property of R

nonsmooth Newton methods, and no stronger than conditions for sequential quadratic programming for optimization.

In conclusion, we note that weak regularity subsumes (2.13), and hence, the R_0-property and 2-regularity are the same under weak regularity. At the same time, weak regularity (and even the nonsingularity of $F'(\bar{x}, \bar{\mu}_I)$), do not subsume the R_0-property, as illustrated by Example 3.2, and 2-regularity does not imply weak regularity, which is illustrated by Example 2.2. On the other hand, according to Proposition 2.1, under SONC and symmetry of $\frac{\partial \Psi}{\partial x}(\bar{x}, \bar{\mu})$, the R_0-property is equivalent to SMFCQ combined with SOSC, and hence, according to Proposition 4.2, the R_0-property subsumes weak regularity in this case (this fact will be important for our comparison with sequential quadratic programming in Section 5). Finally, 2-regularity and weak regularity are both implied by BD-regularity of R, and are strictly weaker.

5 Comparison with Other Newton-Type Methods

In this section, we compare our approach with the generalized Newton method (GNM) for solving nonsmooth reformulations of KKT systems, and with sequential quadratic programming methods (SQP). We shall also comment on the applicability to KKT systems of the method proposed in [12] for complementarity problems. Our focus would be on local convergence properties. Globalization issues will be discussed very briefly at the end of this section.

The first observation is that, in principle, GNM and SQP do not require identification of active constraints. But this difference with our approach is essentially methodological, and we believe it does not result in any concrete advantages/disadvantages for the following reasons. All these methods are local, i.e., they require a starting point sufficiently close to $(\bar{x}, \bar{\mu})$. Furthermore, the regularity assumptions needed for convergence of GNM or SQP (see below) imply the error bound (2.2). Therefore, identification of active constraints comes in this setting at no price, at least as a matter of principle.

For nonsmooth methods, we shall consider the reformulation

$$R(x,\, \mu) = 0, \tag{5.1}$$

where R is defined by (2.1). The GNM for (2.1) is given by

$$(x^{k+1},\, \mu^{k+1}) = (x^k,\, \mu^k) - (\Lambda_k)^{-1} R(x^k,\, \mu^k), \quad \Lambda_k \in \partial_B R(x^k,\, \mu^k). \tag{5.2}$$

We refer the reader to [19, 20, 24, 26] for properties of this type of algorithms. For the stated GNM, the condition required for local convergence is BD-regularity of R at $(\bar{x},\, \bar{\mu})$. We note that using other subdifferentials (e.g., the Clarke subdifferential) or other reformulations (e.g., based on the Fischer-Burmeister function [9, 7]) leads to even stronger regularity conditions [25].

Note that our approach is completely justified under the assumptions of 2-regularity of S at $(\bar{x},\, \bar{\mu})$ and weak regularity, and these conditions are strictly weaker than BD-regularity of R at $(\bar{x},\, \bar{\mu})$ (equivalently, quasi-regularity). In fact, even stronger assertions are established in Sections 2 and 4. We next give a more detailed comparison.

To simplify the comparison, let $N = \emptyset$ (note that everything can be reduced to this setting in any case, via identification of I). Then, locally, (5.1) can be re-written in the form

$$R_{J_1,\, J_2}(x,\, \mu_I) = 0 \tag{5.3}$$

with

$$R_{J_1,\, J_2} : \mathbf{R}^n \times \mathbf{R}^{|I|} \to \mathbf{R}^n \times \mathbf{R}^{|I|}, \quad R_{J_1,\, J_2}(x,\, \mu_I) = \left(\begin{array}{c} \Phi(x) - (G_I'(x))^{\mathrm{T}} \mu_I \\ G_{J_1}(x) \\ \mu_{J_2} \\ G_{I_+}(x) \end{array} \right), \tag{5.4}$$

where $(J_1,\, J_2)$ is a partition of I_0 such that

$$\{i \in I_0 \mid \mu_i > G_i(x)\} \subset J_1, \quad \{i \in I_0 \mid \mu_i < G_i(x)\} \subset J_2. \tag{5.5}$$

Of course, J_1 and J_2 depend on $(x,\, \mu_{I_0})$. The set $\partial_B R(x,\, \mu)$ is comprised by matrices $R'_{J_1,\, J_2}(x,\, \mu_I)$ corresponding to all partitions $(J_1,\, J_2)$ of I_0 satisfying (5.5) at a given point $(x,\, \mu_I) \in \mathbf{R}^n \times \mathbf{R}^{|I|}$ (note that for fixed J_1 and J_2, $R_{J_1,\, J_2}$ is smooth). This means that each iteration (5.2) of GNM can be thought of as a usual Newton iteration for smooth equation (5.3) with the appropriate choice of J_1 and J_2. Of course, in GNM partitions $(J_1,\, J_2)$ can change along the iterations, and the entire process is not the usual Newton method for some smooth equation. But in some sense, this is probably a *disadvantage* rather than an advantage of this approach: it leads to obvious complications and also to stronger regularity conditions, as discussed next.

Namely, let us fix an *arbitrary* partition $(J_1,\, J_2)$ of I_0. Of course, $(\bar{x},\, \bar{\mu}_I)$ is still a solution of (5.3). This solution can be found by the usual Newton method, provided $R'_{J_1,\, J_2}(\bar{x},\, \bar{\mu}_I)$ is nonsingular for the given J_1 and J_2. At the same time, to justify the local superlinear convergence of GNM for (5.1), one has to assume that R is BD-regular at $(\bar{x},\, \bar{\mu}_I)$, i.e.,

$R'_{J_1,\,J_2}(\bar{x},\,\bar{\mu}_I)$ is nonsingular for *all* partitions $(J_1,\,J_2)$ of I_0. Roughly speaking, this is the consequence of the fact that the variables x and μ in (5.5) are "decoupled". For this reason, all partitions of I_0 are locally possible (or "feasible"). Hence, in GNM the sequence of partitions $(J_1,\,J_2)$ of I_0 can have an arbitrary partition as its "limit point". By the way, the above consideration also makes it clear why BD-regularity of R at $(\bar{x},\,\bar{\mu}_I)$ is equivalent to quasi-regularity, as claimed in Section 2 (recall (2.4)).

Now observe that system (5.3) is a particular case of (4.3) corresponding to $A = \mathrm{diag}\{a_i,\, i \in I_0\}$, $a_i = 1,\, i \in J_1$, $a_i = 0,\, i \in J_2$, $B = E - A$. With this choice, $R_{J_1,\,J_2} = F_{A,\,B}$. Our approach permits "less special" matrices A and B, and those matrices do not "oscillate" as in GNM. This leads to weaker regularity assumptions.

Next, we comment on the method developed in [12] for solving complementarity problems, including mixed complementarity [12, Section 5.2]. Since KKT systems are a special case of mixed complementarity, the ideas of [12] are applicable here. The method of [12] is based on reformulating the complementarity problem as a system of smooth degenerate equations, "regularizing" it by a certain nonsmooth term, and then applying a special GNM to the resulting nonsmooth equation. Without going into formal analysis, we believe that applying the approach of [12] in the present setting would require for convergence an assumption different from BD-regularity of R at $(\bar{x},\bar{\mu})$, i.e., an assumption which is not weaker and not stronger. (This should be similar to the way the regularity conditions compare in the case of the nonlinear complementarity problem, see [12, Section 4].) Furthermore, we believe that the resulting assumption would be stronger than weak regularity, required for the method proposed here. Essentially, the latter is due to the fact that the iteration of the method derived for KKT from the approach of [12] should be possible to relate to our framework here by choosing some special matrices A and B, similar to what have been done above for GNM. Note finally that our method here is rather different from the idea of [12], which relies on reformulation and regularization. The mapping (4.4) is not related to any reformulation of the KKT system.

Finally, we compare our approach with SQP methods. We shall focus essentially on the optimization case: we assume that $\frac{\partial\Psi}{\partial x}(\bar{x},\,\bar{\mu})$ is symmetric, and SONC is satisfied. Note that we compare with the classical SQP scheme only, where each iteration consists of solving one quadratic programming subproblem: $x^{k+1} = x^k + d^k$, where $d^k \in \mathbf{R}^n$ is a solution of

$$\begin{array}{ll} \text{minimize} & \langle \Phi(x^k),\, d\rangle + \frac{1}{2}\left\langle \frac{\partial\Psi}{\partial x}(x^k,\,\mu^k)d,\, d\right\rangle \\ \text{subject to} & G(x^k) + G'(x^k)d \geq 0, \end{array} \tag{5.6}$$

and μ^{k+1} is set to be a multiplier associated with the solution d^k. We do not consider here any nontrivial modifications of SQP (e.g., [30, 10, 27]), which require more than solving a quadratic program at each iteration. We also do not consider special cases, such as convexity or even linearity of constraints. At the expense of those additional assumptions and/or complex modifications of the basic SQP idea, some local conditions for convergence can be relaxed (for example, uniqueness of the multiplier). We refer the reader to [6, 10] for some discussions and references.

To our knowledge, the weakest condition guaranteeing local superlinear convergence of the SQP method based on (5.6) is SMFCQ combined with SOSC [2]. By Proposition 2.1, this is equivalent to the R_0-property. As mentioned in Section 4, 2-regularity combined with weak regularity is equivalent to the R_0-property as well. Hence, our approach is completely justified under the same conditions as those needed for the SQP methods. This is quite remarkable, taking into account that our method is QP-free, i.e., it is based on solving linear systems instead of the more costly QP problems. Normally, QP-free methods require strictly stronger local assumptions than those needed for SQP (e.g., see [25, 4], and our discussion above of GNM).

Finally, the SQP method can be related to our approach not only in terms of the associated regularity conditions, but also in terms of the iteration update formula, similarly to GNM. By the KKT conditions for (5.6), $(x^{k+1},\, \mu^{k+1})$ satisfy the following system:

$$\begin{array}{l}\frac{\partial \Psi}{\partial x}(x^k,\, \mu^k)(x^{k+1}-x^k) - (G'(x^k))^{\mathrm{T}}\mu^{k+1} = -\Phi(x^k),\\ \mu^{k+1} \geq 0,\ G(x^k)+G'(x^k)(x^{k+1}-x^k) \geq 0,\ \langle \mu^{k+1},\, G(x^k)+G'(x^k)(x^{k+1}-x^k)\rangle = 0.\end{array} \qquad (5.7)$$

Let $N = \emptyset$, and assume that $\{(x^k,\, \mu^k)\}$ converges to $(\bar{x},\, \bar{\mu})$. Then it is evident from (5.7) that for k large enough there exists a partition $(J_1,\, J_2)$ of I_0 such that

$$\begin{array}{l}\frac{\partial \Psi}{\partial x}(x^k,\, \mu_I^k)(x^{k+1}-x^k) - (G_I'(x^k))^{\mathrm{T}}\mu_I^{k+1} = -\Phi(x^k),\\ G_{J_1}'(x^k)(x^{k+1}-x^k) = -G_{J_1}(x^k),\\ \mu_{J_2}^{k+1} = 0,\\ G_{I_+}'(x^k)(x^{k+1}-x^k) = -G_{I_+}(x^k),\end{array}$$

which is the usual Newton iteration for the equation (5.3), (5.4). Of course, here partitions $(J_1,\, J_2)$ of I_0 are defined differently from GNM. As in GNM, these partitions can change along the iterations, but the possible limit points of the sequence of partitions are "less arbitrary" here. This leads to weaker regularity conditions.

To conclude the discussion, we briefly comment on globalization of the local algorithms mentioned above. In the optimization case, SQP is certainly the method for which there exist the most natural (and easy to implement) globalization strategies [1]. In the case of general KKT systems (e.g., associated with asymmetric variational inequalities), in our opinion, the situation is not as clear. It is generally agreed that the easiest to globalize is GNM applied to the Fischer-Burmeister reformulation of the system. This has to do with the fact that the squared 2-norm of this reformulation is continuously differentiable, which makes it a good choice for the merit function in a linesearch procedure [15, 16]. However, there is no guarantee that globally the Newton direction even exists. The best in this sense seems to be the damped modified Gauss-Newton method for the Fischer-Burmeister reformulation [15, 16], which is globally well-defined. Other known globalization strategies are of the hybrid type: if the Newton direction does not exist or does not yield sufficient descent for the merit function, the globalization algorithm resorts to the steepest descent direction (e.g., [8, Algorithm 1]). As is well explained in [8, Section 4], in such hybrid globalization frameworks one can use essentially any local algorithm with fast convergence. In particular, the method proposed here can be used as the local algorithm within this hybrid strategy. Generally speaking, as a method which relies on identification of active constraints, Algorithm 4.1 is not meant to be independent. It should be incorporated within some globally convergent scheme, for the purpose of achieving fast local convergence.

Acknowledgment. We would like to thank Francisco Facchinei whose comments on our paper [14] led us to consider the error bound presented here.

References

[1] D.P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmont, Massachusetts, 1995.

[2] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[3] J.F. Bonnans and A. Sulem. Pseudopower expansion of solutions of generalized equations and constrained optimization problems. *Mathematical Programming*, 70:123–148, 1995.

[4] F. Facchinei, A. Fischer, and C. Kanzow. Regularity properties of a new equation reformulation of variational inequalities. *SIAM Journal on Optimization*, 8:850–869, 1998.

[5] F. Facchinei, A. Fischer, and C. Kanzow. On the accurate identification of active constraints. *SIAM Journal on Optimization*, 9:14–32, 1999.

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

[7] F. Facchinei and J. Soares. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal on Optimization*, 7:225–247, 1997.

[8] M.C. Ferris, C. Kanzow, and T.S. Munson. Feasible descent algorithms for mixed complementarity problems. *Mathematical Programming*, 86:475–497, 1999.

[9] A. Fischer. A special Newton-type optimization method. *Optimization*, 24:269–284, 1992.

[10] A. Fischer. Modified Wilson's method for nonlinear programs with nonunique multipliers. *Mathematics of Operations Research*, 24:699–727, 1999.

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

[12] A.F. Izmailov and M.V. Solodov. Superlinearly convergent algorithms for solving singular equations and smooth reformulations of complementarity problems, 2000. *SIAM J. on Optimization*, to appear.

[13] A.F. Izmailov and M.V. Solodov. The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions, 2000. *Mathematics of Operations Research*, to appear.

[14] A.F. Izmailov and M.V. Solodov. Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. *Mathematical Programming*, 89:413–435, 2001.

[15] H. Jiang. Global convergence analysis of the generalized Newton and Gauss-Newton methods of the Fischer-Burmeister equation for the complementarity problem. *Mathematics of Operations Research*, 24:529–543, 1999.

[16] H. Jiang and D. Ralph. Global and local superlinear convergence analysis of Newton-type methods for semismooth equations with smooth least squares. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 181–210. Kluwer Academic Publishers, 1999.

[17] C. Kanzow. Some equation-based methods for the nonlinear complementarity problem. *Optimization Methods and Software*, 3:327–340, 1994.

[18] M. Kojima. Strongly stable stationary solutions in nonlinear programs. In S.M. Robinson, editor, *Analysis and Computation of Fixed Points*, pages 93–138. Academic Press, New York, 1980.

[19] B. Kummer. Newton's method for nondifferentiable functions. In J. Guddat et al., editor, *Advances in Mathematical Optimization*, volume 45, pages 114–125. Akademie-Verlag, Berlin, 1988.

[20] B. Kummer. Newton's method based on generalized derivatives for nonsmooth functions. In W. Oettli and D. Pallaschke, editors, *Advances in Optimization*, pages 171–194. Springer–Verlag, Berlin, 1992.

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

[22] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[23] J.-S. Pang. Private communication, January 2001.

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

[25] L. Qi and H. Jiang. Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. *Mathematics of Operations Research*, 22:301–325, 1997.

[26] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58:353–367, 1993.

[27] L. Qi and Z. Wei. On the constant linear dependence condition and its application to an SQP method. *SIAM Journal on Optimization*, 10:963–981, 2000.

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

[29] P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 89:17–37, 1996.

[30] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Computational Optimization and Applications*, 11:253–275, 1998.