Preprint-version 2006

Newton methods for stationary points: an elementary view of regularity conditions and solution schemes

Diethard Klatte[1] Bernd Kummer [2]

Submitted 8 July 2005, Revised version 8 March 2006

Abstract. In this paper, we give an elementary view of Newton-type methods and related regularity conditions for a special class of nonsmooth equations arising from necessary optimality criteria for standard nonlinear programs. Different types of linearizations and parameterizations of these equations lead to different iteration schemes, where any abstract calculus of generalized derivatives for nonsmooth mappings is avoided. Based on a general local convergence result on (perturbed) Newton methods for solving Lipschitzian equations, we focus on characterizations which are explicitly given in terms of the original functions and assigned quadratic problems for our special setting. We are particularly interested in certain parameterized Newton equations and in regularity conditions which are weaker than strong regularity.

Key words. Generalized Newton methods, stationary solutions, nonlinear programs, regularity, particular realizations of Newton equations.

AMS classification. 90C31, 49J40, 49J52, 26E25.

1 Introduction

We consider the nonlinear optimization problem

$$\min\ f(x) \text{ subject to } g_i(x) \le 0,\ i = 1, \dots, m; \quad f, g_i \in C^2(\mathbb{R}^n, \mathbb{R}), \tag{1.1}$$

and we suppose throughout that $D^2 f$, $D^2 g_i$ are locally Lipschitzian (briefly $f, g_i \in C^{2,1}$). Necessary optimality conditions will be used in terms of *Kojima's function* (see [12]) $\Phi : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ which has the components

$$\begin{array}{rclcrcl} \Phi_1(x, y) & = & Df(x) & + & \sum_{i=1}^m y_i^+ Dg_i(x), & & y_i^+ = \max\{0, y_i\}, \\ \Phi_{2,i}(x, y) & = & g_i(x) & - & y_i^-, & & y_i^- = \min\{0, y_i\}. \end{array} \tag{1.2}$$

Then the zeros of Φ are related to the KKT points of (1.1) via the transformations

$$\begin{array}{lcl} (x, y) \in \Phi^{-1}(0) & \Rightarrow & (x, u) = (x, y + g(x)) \text{ is a KKT-point} \\ (x, u) \text{ is a KKT-point} & \Rightarrow & (x, y) = (x, u + g(x)) \in \Phi^{-1}(0). \end{array} \tag{1.3}$$

Obviously, primal (stationary) solutions are the same in both descriptions, dual ones differ only in the case $g_i(x) < 0$ which gives $u_i = 0$ but $y_i = g_i(x)$. The first equation in (1.2) may also be written as $D_x L(x, y^+) = 0$, where

$$L(x, y) := f(x) + \textstyle\sum_{i=1}^m y_i g_i(x)$$

denotes as usual the *Lagrange function* associated with (1.1). Further, it can be easily seen that for $(a, b) \in \mathbb{R}^{n+m}$,

$$\begin{array}{l} (x, y) \in \Phi^{-1}(a, b) \text{ iff both } \quad y^- = g(x) - b \text{ and} \\ (x, y^+) \text{ is a KKT-point of the perturbed problem} \end{array} \tag{1.4}$$

$$\min\ f(x) - \langle a, x \rangle \text{ subject to } g(x) \le b. \tag{1.5}$$

[1]Address: Institut für Operations Research,Universität Zürich, Moussonstrasse 15, CH-8044 Zürich, Switzerland. E–Mail: klatte@ior.unizh.ch

[2]Address: Institut für Mathematik, Humboldt–Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. E–Mail: kummer@mathematik.hu-berlin.de

In the present paper, we discuss several versions of Newton-type methods for solving the key system

$$\Phi(x, y) = 0. \tag{1.6}$$

This is a subject of active research in the last two decades (we refer, e.g., to [8, 13, 14, 15, 21, 23, 18, 10, 5]). Our purpose is to present an elementary view of different standard and perturbed Newton schemes and their local convergence analysis, avoiding the use of prerequisites from nonsmooth analysis. Instead, we strictly utilize the special structure of the equations (1.2). Further, we also intend to give (as much as possible) a self-contained presentation of the material. The reader may consult [10, 17, 18] for further details.

In section 2, we start with local convergence analysis of Newton's method for solving (1.6) with an arbitrary locally Lipschitzian Φ and for a general type of approximation of Φ. In the present form, the result is new and extends Theorem 10.7 in [10]. Then, in section 3, we clarify how to guarantee the conditions in the setting of (1.2). The involved non-differentiable functions do not make serious problems, since non-smoothness is essentially that of the absolute value function. Our focus is on subsets of the known Clarke derivative, this allows relatively simple Newton steps and sometimes regularity conditions weaker than strong regularity in Robinson's [22] sense. For comparison, regularity in terms of directional derivatives is discussed.

The investigations in the sections 4 and 5 are motivated by the following observation. If one applies Newton's method directly to (1.2) and $y_i^k > 0$ holds at the iteration point (x^k, y^k) then the Newton system involves the linear equation (assigned to $\Phi_{2,i}$)

$$g_i(x^k) + Dg_i(x^k)^{\mathsf{T}}(x - x^k) = 0 \tag{1.7}$$

since $\frac{\partial \Phi_{2,i}}{\partial y_i}$ and y_i^- vanish at y_i^k. Thus, if the number of such i is higher than n, system (1.7) degenerates and (usually) the method fails to work.

To avoid this effect (which may appear if (x^k, y^k) is not close enough to a "regular" solution), we perturb system (1.6) by adding some appropriate small function or (alternatively) deform the matrix of the Newton equations only. In any case, this is some kind of regularization. As a first idea, one could add a term εy_i to function $\Phi_{2,i}(x, y)$ in the above situation. We shall consider, however, other regularizations which can be handled surprisingly simple.

The meaning of such regularizations and the permitted amount (for ensuring still quadratic convergence of the whole method) will be studied in the sections 4 and 5 under more and less general settings, respectively. In particular, we recall that some of these regularizations describe just the stationary points in penalty and barrier methods, assigned to (1.1). So the content of our regularizations becomes evident. On the other hand, the both classical methods turn out to be specific regularization methods with respect to the equations (1.7) for $y_i^k > 0$, only. Finally, we interpret the related Newton steps in terms of SQP-methods.

2 Newton's method

In this section we recall basic facts on generalized Newton methods from [10] and extend convergence estimates for standard Newton schemes to a perturbed one. Throughout Section 2, we choose a setting more general than in the rest of the paper and consider the equation

$$\Phi(z) = 0,$$

where $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitzian with rank (i.e., with Lipschitz constant) L_Φ near a zero z^* of Φ. Let for any z in some neighborhood Ω of z^*, $D\Phi(z) : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a multivalued mapping (closed or not) satisfying

$$D\Phi(z)(\zeta) \neq \emptyset \quad \text{and} \quad D\Phi(z)(0) = \{0\}.$$

Suppose that there is some $K_R > 0$ such that for all z in some ball $B(z^*, R)$, the following *condition for the approximation* is satisfied:

$$\text{(CA)} \qquad \Phi(z) - \Phi(z^*) + D\Phi(z)(z^* - z) \subset B(0, K_R\|z - z^*\|^2). \tag{2.1}$$

Suppose, in addition, an *injectivity (regularity) condition* near z^*:

$$\begin{array}{l} \text{(CI)} \quad \inf \{ \ \|\phi\| \ : \ \phi \in D\Phi(z)(\zeta) \ \} \ \geq \ K_I \ \|\zeta\| \\ \text{for some } K_I > 0 \text{ and all } z \text{ in some ball } B(z^*, \delta). \end{array} \tag{2.2}$$

Then $D\Phi(z)(\cdot)$ may be applied in the following

Standard Newton scheme: Given z^k find $z^{k+1} = z$ such that

$$0 \in \Phi(z^k) + D\Phi(z^k)(z - z^k). \tag{2.3}$$

Then, if z^{k+1} exist, the known estimates of the case of C^1 equations are valid, cf. Theorem 10.7 in [10], namely

$$\begin{array}{l} \|z^{k+1} - z^*\| \ \leq \ K_I^{-1} K_R \ \|z^k - z^*\|^2 \ \text{ and } \ \|z^{k+1} - z^*\| \leq \frac{1}{2}\|z^k - z^*\| \\ \text{whenever } \ \|z^0 - z^*\| \ \leq \ r = \min\{R, \ \delta, \ \frac{1}{2}K_R^{-1} \ K_I\}. \end{array} \tag{2.4}$$

Remark 1

1. It suffices that (2.1) and (2.2) are satisfied for the iteration points z^k.
2. The existence of a solution z^{k+1} for (2.3) is evident if

$$D\Phi(z)(\zeta) = S(z)\zeta := \{A\zeta \ : \ A \in S(z)\} \tag{2.5}$$

   holds with a set of matrices $S(z)$ of appropriate dimension (which will be satisfied for the generalized derivatives considered in the cases 2 ... 5 of (3.3), (3.4) under intrinsic assumptions, cf. §3.1 below). Then injectivity (2.2) ensures regularity of all $A \in S(z)$ and (2.3) can be solved via $\Phi(z^k) + A(z - z^k) = 0$ with any $A \in D\Phi(z^k)$.
3. If (CA) holds with Clarke's generalized Jacobian $S(z) = \partial\Phi(z)$ in (2.5), then Φ is called *strongly semismooth* at z^*.
4. For the importance of (CA) and (CI) in general, relations to semi-smoothness [19, 21] and approximate solutions of (2.3), see [14, 15, 10].
5. If $D\Phi$ satisfies (CA) and (CI) so also all selection functions $s\Phi(z) \in D\Phi(z)$ with $s\Phi(z)(\zeta) \in D\Phi(z)(\zeta)$ satisfy these conditions.

Now we show that the estimate (2.4) can be extended to a *perturbed Newton scheme* which we shall need for subsequent perturbations of Φ.

Theorem 1 (Perturbations of $D\Phi$) *Suppose (2.1) and (2.2) and consider a perturbed system*

$$0 \in \Phi(z^k) + D\Phi(z^k)(z - z^k) + A^k(z - z^k), \quad z^{k+1} := z \tag{2.6}$$

where A^k is supposed to be a matrix of suitable order such that

$$\|A^k\| \leq C \ \|\Phi(z^k)\| \textit{ for some constant } C. \tag{2.7}$$

Then, the estimates (2.4) are again true, only K_R and K_I must be replaced by

$$K_R' = K_R + CL_\Phi \textit{ and } K_I' = \tfrac{1}{2}K_I,$$

and the radius

$$r \le \min\{R,\ \delta,\ \tfrac{1}{2}(K_R')^{-1}\ K_I'\} \tag{2.8}$$

from (2.4) has, in addition, to satisfy

$$r \le \frac{K_I}{2CL_\Phi} \text{ where } L_\Phi \text{ is a Lipschitz rank of } \Phi \text{ on } B(z^*, R). \tag{2.9}$$

Supplement: Having (2.7), (2.8) and (2.9), then it holds

$$\|\Phi(z^{k+1})\| \ \le\ \|\Phi(z^k)\| \quad \text{if } \|\Phi(z^k)\| \le \frac{K_I^3}{8L_\Phi K_R'}, \tag{2.10}$$

and for the special form (2.5) of $D\Phi(z^k)$, the solutions z^{k+1} of (2.6) exists and can be obtained by solving $\Phi(z^k) + (A + A^k)(z - z^k) = 0$ with any $A \in D\Phi(z^k)$. ◇

Proof. We have from (2.1) and (2.2), for $z = z^k$ with $\|z - z^*\| \le r = \min\{R,\ \delta,\ \frac{K_I}{2K_R}\}$,

$$D\Phi(z)(z^* - z) \subset \Phi(z^*) - \Phi(z) + B(0,\ K_R\ \|z - z^*\|^2), \tag{2.11}$$

and $\phi \in D\Phi(z)(z^* - z)$ yields $\|\phi\| \ge\ K_I\ \|z - z^*\|$. Thus,

$$\|\Phi(z)\|\ \ge\ K_I\ \|z - z^*\| - K_R\ \|z - z^*\|^2 = \|z - z^*\|\ (K_I - K_R\|z - z^*\|)$$

which ensures that

$$\|\Phi(z)\| \ge\ \tfrac{1}{2}K_I\|z - z^*\| \quad \text{if } \|z - z^*\| \le \frac{K_I}{2K_R}. \tag{2.12}$$

The latter holds due to (2.8). Since Φ is locally Lipschitz, this yields for related z,

$$\tfrac{1}{2}K_I\|z^k - z^*\| \le \|\Phi(z^k)\| \le L_\Phi\|z^k - z^*\|. \tag{2.13}$$

So (2.7) implies that

$$\|A^k\| \le CL_\Phi\|z^k - z^*\|.$$

With the new derivative $D\Phi(z^k)(\zeta) + A^k\zeta$ at z^k and the new constant $K_R' = K_R + CL_\Phi$, this ensures in (2.1) the required estimate. In (2.2) we obtain for $\phi \in D\Phi(z)(z^* - z) + A^k(z^* - z)$,

$$\|\phi\| \ge K_I\|z - z^*\| - \|A^k\|\ \|z - z^*\| \ge K_I\|z - z^*\| - CL_\Phi\|z^k - z^*\|^2 \ge \tfrac{1}{2}K_I\ \|z - z^*\|,$$

whenever

$$\|z - z^*\| \le \frac{K_I}{2CL_\Phi}. \tag{2.14}$$

Hence (2.8) and (2.9) guarantee all needed estimates to apply (2.4) with the new constants. Further, injectivity ensures the existence of z^{k+1} under (2.5) via regularity of all matrices in $S(z^k) + A^k$.

Monotonicity of the error: Having (2.8) and (2.9) then, applying $K_I' = \frac{1}{2}K_I$ and (2.4) with the new constants, we obtain

$$\|z^{k+1} - z^*\|\ \le 2K_I^{-1}K_R'\|z^k - z^*\|^2. \tag{2.15}$$

With $\|\Phi(z^{k+1})\| \le L_\Phi\|z^{k+1} - z^*\|$ and $\|z^k - z^*\| \le 2K_I^{-1}\|\Phi(z^k)\|$ from (2.13), it follows

$$\|\Phi(z^{k+1})\| \le L_\Phi\|z^{k+1} - z^*\| \le 2L_\Phi K_I^{-1}K_R'\|z^k - z^*\|^2 \le 8L_\Phi K_I^{-3}K_R'\ \|\Phi(z^k)\|^2. \tag{2.16}$$

Hence (2.10) is valid. □

Remark 2 By the proof one easily sees that Theorem 1 holds more general. Finite dimensionality and some particular structure of $D\Phi$ has been used only for the existence of z^{k+1}. All other arguments were norm-estimates which hold in Banach spaces, too (then $A^k : X \to Y$ is a linear operator).

3 Representations of $D\Phi$ and the injectivity condition (CI)

Now, as in the rest of the paper, let Φ be the Kojima function (1.2). Following [10] where the details of the subsequent statements can be found, the mapping Φ can be written as a product $\Phi(x,y) = M(x)N(y)$ where N and M have the form

$$N(y) = (1, \ \ y_1^+, ..., y_m^+, \ \ y_1^-, ..., y_m^-)^\mathsf{T} \in \mathbb{R}^{1+2m}$$

$$M(x) = \begin{pmatrix} Df(x) & Dg_1(x) & ... & Dg_i(x) & ... & Dg_m(x) & 0 & ... & 0 & ... & 0 \\ g_i(x) & 0 & ... & 0 & ... & 0 & 0 & ... & -1 & ... & 0 \end{pmatrix} \tag{3.1}$$

with $i = 1, ..., m$ and -1 at position i in the last block.

3.1 The functions $D\Phi$

Since Φ is a nonsmooth function, the use of some generalized derivative in Newton's method is required. However, nonsmoothness is only implied by the elementary piecewise linear map N which is basically defined by the components

$$\mu(y_i) = (y_i^+, y_i^-) = (y_i^+, y_i - y_i^+) = \tfrac{1}{2}(y_i + |y_i|, y_i - |y_i|), \quad i = 1, ..., m. \tag{3.2}$$

So, our discussions on generalized derivatives will be reduced to the question of how to define a suitable derivative of the *absolute value function* $s \in \mathbb{R} \mapsto \alpha(s) = |s|$ at the origin in direction σ. We consider the following five possibilities:

$$\begin{array}{ll} 1. & D\alpha(0)(\sigma) = \lim_{\lambda \downarrow 0} \lambda^{-1}\alpha(\lambda\sigma), \\ 2. & D\alpha(0)(\sigma) = \{\lambda\sigma : \ -1 \le \lambda \le 1\}, \\ 3. & D\alpha(0)(\sigma) = \{-\sigma, \sigma\}, \\ 4. & D\alpha(0)(\sigma) = \ \ \sigma, \\ 5. & D\alpha(0)(\sigma) = -\sigma, \end{array} \tag{3.3}$$

while $D\alpha(s)(\sigma) = \alpha'(s)\sigma$ for $s \neq 0$ in all cases. The above listed derivatives immediately carry over by component-wise definition to the corresponding generalized directional derivatives $D\mu(y_i)(\sigma)$ and $DN(y)(v)$ which are set-valued in the cases 2 and 3.

Given one of the five types of derivatives, we now *define* a (possibly set-valued) generalized derivative of Φ by the usual product rule

$$D\Phi(x,y)(u,v) := [DM(x)u]N(y) + M(x)[DN(y)(v)], \tag{3.4}$$

where $DM(x)$ denotes the Fréchet derivative of the C^1 mapping M at x.

Similarly, if $\psi = \psi(x,y)$ is any function that can be written with differentiable h as $\psi(x,y) = h(x, y, |y_1|, \dots, |y_m|)$ we "differentiate" it by combining the usual chain rule with (3.3).

Because of $f, g \in C^2$ and the special form of N, $D\Phi$ represents in the cases 1, 2 and 3 certain generalized derivatives which are well-known from the literature, for details of proving their coincidence with (3.4) in the related case, we refer e.g. to [10, Chapters 6,7]:

In *case 1*, $D\Phi$ (similarly for $D\alpha$, $D\mu$ and DN) is the usual *directional derivative* of Φ at (x,y) in direction (u,v), which coincides in our setting with the *contingent derivative* [1]. In *case 2*, $D\alpha(s)(\sigma)$ can be interpreted as the *Thibault derivative* [24, 25] which coincides with $[\partial\alpha(s)]\sigma$, where $\partial\alpha(s)$ is *Clarke's subdifferential* [3, 4]; these types of derivatives carry over to $D\mu$, DN and $D\Phi$ and again coincide. The derivative in case 3 results from applying the so-called (see e.g. [5]) B-subdifferential to a given direction, for brevity, we call it *B-derivative*.

In the cases 2...5, condition (2.5) is satisfied, which was essential in Theorem 1 for ensuring the existence of a solution z^{k+1} of the k-th Newton "equation".

3.2 The injectivity condition (CI) at a point and near the solution

For the Kojima function Φ (1.2), the assumption $f, g_i \in C^{2,1}$ and the simple structure of N (and Φ) ensure at $z^* = (x^*, y^*) \in \Phi^{-1}(0)$ that for all $z = (x, y)$ in some ball $B(z^*, R)$ and all "derivatives" $D\Phi$ given by (3.3) (3.4), there is some $K_R > 0$ such that the *approximation condition* (CA) is satisfied.

In contrast, the meaning of the *injectivity condition* (CI) for the original problem (1.1) depends not only on the concrete form of Φ, but also on the applied derivative.

Given some $z^* = (x^*, y^*) \in \Phi^{-1}(0)$, we put

$$I^+ = \{i : y_i^* > 0\},\ I^- = \{i : y_i^* < 0\} \text{ and } I^0 = \{i : y_i^* = 0\}.$$

Further, let

$$Y^* = \{y \mid \Phi(x^*, y) = 0\}$$

denote the set of multipliers associated with x^*.

Regular matrices

In the cases 2, ..., 5 considered in (3.3)-(3.4) above, the injectivity condition (2.2) requires specific regularity properties of certain matrices $G(x, y, p)$, where $p_i \in \{0, 1\}$ in each case.

These matrices can be obtained from the product rule (3.4) and have for given (x, y, p) the following structure:

$$G(x,y,p) = \begin{pmatrix} D_x^2 L(x, y^+) & p_1 Dg_1(x) & \dots & p_m Dg_m(x) \\ Dg_1(x)^\mathsf{T} & -(1-p_1) & & \\ \vdots & & \ddots & \\ Dg_m(x)^\mathsf{T} & & & -(1-p_m) \end{pmatrix}. \tag{3.5}$$

To see this, let us repeat the main arguments from [10]. Taking case 2 as the basic concept, we have for $\mu_i(y_i) = (y_i^+, y_i^-)$ in direction v_i that

$$D\mu_i(y_i)(v_i) = \{(p_i v_i, (1-p_i)v_i) \mid 0 \le p_i \le 1\} \quad \text{if } y_i = 0$$

and $D\mu_i(y_i)(v_i) = \{(v_i, 0)\}$ if $y_i > 0$, but $D\mu_i(y_i)(v_i) = \{(0, v_i)\}$ if $y_i < 0$. Define

$$\mathcal{R}(y) := \{p = (p_1, \dots, p_m) \in [0,1]^m \mid p_i = 1 \text{ if } y_i > 0,\ p_i = 0 \text{ if } y_i < 0\}.$$

Then we immediately obtain in case 2 that

$$DN(y)(v) = \{(0\ ,\ \dots, p_i v_i, \dots,\ \dots, (1-p_i)v_i, \dots)^\mathsf{T} \mid p \in \mathcal{R}(y)\} \tag{3.6}$$

for a given direction v, hence the second term in the product rule (3.4) becomes

$$M(x)[DN(y)(v)] = \left\{ \begin{pmatrix} \sum_{i=1}^m p_i v_i Dg_i(x) \\ -\sum_{i=1}^m (1-p_i) v_i \end{pmatrix} \middle|\ p \in \mathcal{R}(y) \right\}.$$

Further, since $DM(x)$ is the standard Fréchet derivative at x, the first term in the product rule (3.4) is easily computed to be

$$[DM(x)u]N(y) = \begin{pmatrix} D^2 f(x)u + \sum_{i=1}^m y_i^+ D^2 g_i(x)u \\ Dg(x)^\mathsf{T} u \end{pmatrix}.$$

Hence, in case 2 (Thibault/Clarke derivative) one has

$$[DM(x)u]N(y) + M(x)[DN(y)(v)] = \left\{ G(x,y,p) \begin{pmatrix} u \\ v \end{pmatrix} \,\middle|\, p \in \mathcal{R}(y) \right\} \tag{3.7}$$

and injectivity (CI) then means that $\det G(x,y,p) \neq 0$ holds for each (x,y) near (x^*,y^*) and all $p \in \mathcal{R}(y)$. Clearly, this is true for small distance $\|(x,y)-(x^*,y^*)\|$ whenever $\det G(x^*,y^*,p) \neq 0$ for all $p \in \mathcal{R}(y^*)$. So the "neighborhood condition" (CI) is satisfied if it holds true at the solution.

Moreover, the set $\{G(x^*,y^*,p) \,|\, p \in \mathcal{R}(y^*)\}$ is arc-wise connected, hence $\det G(x^*,y^*,p)$ has under (CI) the same sign for these matrices. Because each p_i appears in exactly one column and this in a affine–linear manner, the function $p \mapsto G(x^*,y^*,p)$ is affine–linear in each p_i, too. So (by induction arguments) it suffices only to check whether all determinants $\det G(x^*,y^*,p)$ for $p \in \mathcal{R}(y^*)$ and $p_i \in \{0,1\}$ have the same non-vanishing sign.

In the cases 3-5, the formulas (3.6) and (3.7) remain true after restricting $\mathcal{R}(y)$ to such p satisfying for $y_i = 0$ the special settings $p_i \in \{0,1\}$ (case 3), $p_i = 1$ (case 4) and $p_i = 0$ (case 5), respectively. This immediately leads to the following characterizations of *injectivity (2.2) at a point* $z = (x,y)$. We denote this property by $(CI)_z$.

Case 2 (Thibault/Clarke derivative): The condition (2.2) is equivalent to the requirement that sign $\det G(x^*,y^*,p)$ is constant and not zero for all p satisfying

$$p_i = 1 \text{ if } i \in I^+, \quad p_i \in \{0,1\} \text{ if } i \in I^0, \quad p_i = 0 \text{ if } i \in I^-. \tag{3.8}$$

Case 3 (B-derivative): (2.2) is equivalent to $\det G(x^*,y^*,p) \neq 0$ for all p from (3.8).

Indeed, (2.2) is equivalent to $\det G(x,y,p) \neq 0$, i.e., $(CI)_z$ for $z = (x,y)$ near (x^*,y^*) and all $p \in \mathcal{R}(y)$. Since $\mathcal{R}(y) \subset \mathcal{R}(y^*)$ for y near y^*, so the pointwise condition (often called B-regularity) is sufficient and necessary.

Case 4 ($D\alpha(0) = 1$): $(CI)_z$ is equivalent to $\det G(x,y,p) \neq 0$ for some particular p, namely

$$p_i = 1 \text{ if } y_i \geq 0, \quad p_i = 0 \text{ if } y_i < 0.$$

Case 5 ($D\alpha(0) = -1$): Similarly, $(CI)_z$ is equivalent to $\det G(x,y,p) \neq 0$ for p with

$$p_i = 1 \text{ if } y_i > 0, \quad p_i = 0 \text{ if } y_i \leq 0.$$

In the cases 4 and 5, the considered selections $p(y)$ of $p \in \mathcal{R}(y)$ are not continuous. So the related conditions for $z^* = (x^*,y^*)$ cannot be extended on a neighborhood. On the other hand, the (pointwise) condition of case 3 is, of course, again a sufficient one for ensuring (2.2) in the cases 4 and 5.

Case 1 is omitted here, since the resulting Newton auxiliary problems become linear complementarity problems and the crucial matrix properties become more complicated. For details, we refer to [5] or [10, §7.4.1].

(CI) and quadratic approximations

Depending on $D\Phi$, the injectivity condition (2.2) or the injectivity $(CI)_{z^*}$ at a solution point $z^* = (x^*,y^*) \in \Phi^{-1}(0)$ have also meanings in view of stability of the perturbed problems (1.5) at z^*.

Case 1: For the contingent (directional) derivative of Φ, $(CI)_{z^*}$ characterizes just both the upper regularity (4.2) along with uniqueness of the Lagrange multipliers $Y^* = \{y^*\}$ (strict MFCQ). In addition, violation of $(CI)_{z^*}$ (and hence of (2.2)) is particularly implied by violation of the upper Lipschitz property (4.2) (i). The latter means that there is a non-zero KKT-point (u, v) for the quadratic problem

$$\min_u \ \tfrac{1}{2}u^\mathsf{T} D_x^2 L(x^*, y^+)u \ \text{ s.t. } Dg_i(x^*)^\mathsf{T} u = 0 \ \forall i \in I^+, \quad Dg_i(x^*)^\mathsf{T} u \le 0 \ \forall i \in I^0,$$

for the proof see [10, Cor. 8.18].

Case 2: With the Thibault-derivative or $S(z) = \partial\Phi(z)$ (Clarke's Jacobian [3]) in (2.5), condition (2.2) is equivalent to strong regularity of problem (1.1) at $z^* = (x^*, y^*)$ in Robinson's sense [22], i.e., Φ^{-1} is locally unique and Lipschitz near $(0, z^*)$.

In addition violation of (2.2) means that there is a non-zero KKT-point for some of the quadratic problems

$$Q_J: \quad \min_u \ \tfrac{1}{2}u^\mathsf{T} D_x^2 L(x^*, y^+)u \ \text{ s.t. } Dg_i(x^*)^\mathsf{T} u = 0 \, \forall i \in J, \quad Dg_i(x^*)^\mathsf{T} u \le 0 \, \forall i \in I^0 \setminus J,$$

where $I^+ \subset J \subset I^+ \cup I^0$, cf. [22] or [10, Cor. 8.8]. Similarly, one treats the following cases.

Case 3: For B- derivatives, singularity means that there is a non-zero KKT-point for some of the quadratic problems

$$P_J: \quad \min_u \ \tfrac{1}{2}u^\mathsf{T} D_x^2 L(x^*, y^{*+})u \quad \text{s.t.} \quad Dg_i(x^*)^\mathsf{T} u = 0 \quad \text{if } i \in J,$$

where $I^+ \subset J \subset I^+ \cup I^0$.

Case 4: $D\alpha(0) = 1$: Let us use this "derivative" for differentiating N and computing $D\Phi$ in accordance to (3.4). Considering again the matrix $G(x^*, y^*, p)$ one obtains that $(CI)_{z^*}$ is violated iff there is a non-zero KKT-point (u, v) of the quadratic problem

$$\min_u \ \tfrac{1}{2}u^\mathsf{T} D_x^2 L(x^*, y^{*+})u \quad \text{s.t.} \quad Dg_i(x^*)^\mathsf{T} u = 0 \quad \text{if } i \in I^0 \cup I^+.$$

Case 5: $D\alpha(0) = -1$: Now the related quadratic auxiliary problem has less constraints

$$\min_u \ \tfrac{1}{2}u^\mathsf{T} D_x^2 L(x^*, y^{*+})u \quad \text{s.t.} \quad Dg_i(x^*)^\mathsf{T} u = 0 \quad \text{if } i \in I^+, \tag{3.9}$$

and violation of $(CI)_{z^*}$ is equivalent to the existence of a non-zero KKT-point of (3.9).

Obviously, the existence of a particular KKT point (u, v) with $u = 0$ and $v \ne 0$ means in all cases that $\{Dg_i(x^*),\ i \in I^+ \cup I^0\}$ is linearly dependent, i.e., the linear independence constraint qualification (LICQ) is violated at x^*.

By all means, for the most elementary cases 4 and 5, condition $(CI)_{z^*}$ is weaker than for the cases 2 and 3.

4 Perturbations of the Kojima function

In this section, we study general and special parametrizations Φ^t of the Kojima function Φ and apply this to the standard and perturbed Newton method for solving the original

Kojima system. Given some point x^*, let the following *hypotheses* be satisfied:

$$\begin{array}{ll} (i) & x^* \text{ is a } \textit{local minimizer} \text{ of (1.1) satisfying MFCQ,} \\ (ii) & u^\mathsf{T} D_x^2 L(x^*, y^*) u > 0 \quad \forall y^* \in Y^* \ \forall u \in U^* \setminus \{0\}, \end{array} \tag{4.1}$$

where

$$U^* = \{u|\ Df(x^*)^\mathsf{T} u = 0,\ Dg_i(x^*)^\mathsf{T} u \leq 0 \ \forall\ i : g_i(x^*) = 0\}$$

is the usual *critical cone* at x^* and Y^* denotes the set of all (dual vectors) y satisfying $\Phi(x^*, y) = 0$. By (i), Y^* is nonempty and bounded. Condition (ii) is a standard second-order sufficient optimality condition for (1.1).

Under (4.1) the (multivalued) inverse Φ^{-1} is locally Lipschitzian upper semi–continuous and nonempty-valued in the following sense, cf. [10], Corollary 2.9 and Theorem 8.36.

Lemma 1 *Suppose (4.1). Then there are positive r, c and ρ such that the locally inverse sets $H(a,b) = \{\ (x,y) \in \Phi^{-1}(a,b) \mid \|x - x^*\| \leq r\ \}$ satisfy*

$$\emptyset \neq H(a,b) \ \subset\ (x^*, Y^*) \ +\ c\ \|(a,b)\| B \quad \textit{if} \quad \|(a,b)\| \leq \rho. \tag{4.2}$$

◇

Here and in the rest of the paper, (X, Y) denotes the cartesian product of the sets X and Y, while $(x, Y) = (\{x\}, Y)$. The Lemma ensures persistence and some stable behavior of the KKT-points under small (simple, *canonical*) variations of the initial problem according to (1.5),

$$\min\ f(x) - \langle a, x\rangle \quad \text{subject to} \quad g(x) \leq b.$$

Remark 3 Since Φ is continuous and Y^* is bounded, the mapping H is closed on the ball $B(0, \rho)$. Therefore, if $\Phi^{-1}(a,b)$ is convex or if the component x in $H(a,b)$ is unique then H is trivially closed, convex-valued and uniformly bounded for (a,b) near the origin.

4.1 Parameterized functions Φ^t

We shall consider general and particular perturbations of Φ that arise from nonlinear variations $\psi(x,y,t) \in \mathbb{R}^{n+m}$, where $t \in \mathbb{R}^p$ (some parameter space),

$$\Phi^t(z) = \Phi(z) \ -\ \psi(z,t), \quad z = (x,y). \tag{4.3}$$

We suppose (for particular examples see section 5)

$$\begin{array}{l} \psi \text{ is continuous, and for each } (z,t), \\ \psi(\cdot,t) \text{ is globally Lipschitz with a rank } L(t) \leq C_\psi \|t\|, \\ \psi(z,0) = 0 \text{ and a Lipschitz condition of the type} \\ \|\psi(z,t') - \psi(z,t)\| \leq K\|z\|\ \|t' - t\| \text{ holds true.} \end{array} \tag{4.4}$$

Then $\psi(z,t)$ vanishes for bounded z and $t \to 0$, and

$$\|D_z\psi(\cdot,t)\| := \sup\{\|w\| \mid w \in D_z\psi(.,t)|\} \leq C_\psi\ \|t\|$$

holds for all mentioned generalized derivatives D (Clarke, Thibault, Contingent and B-derivative).

In order to apply the same simple "derivatives" (3.3) to $\psi(.,t)$ we also suppose that ψ can be written as

$$\psi(x,y,t) = h(x,y,|y_1|,...,|y_m|,t) \quad \text{with some } h \in C^{2,1} \tag{4.5}$$

and that we "differentiate" it by combining the usual chain rule with (3.3). By this convention we can use $\psi(z,t)$ for the following general

$$\begin{array}{l}\text{Solution scheme ALG1 :}\\ \text{Given } \; z^k=(x^k,y^k) \; \text{ and some constant } C \text{, choose any } t(k) \text{ such that}\\ \|t(k)\| \le C\,\|\Phi(z^k)\| \quad \text{and find some } z \text{ which satisfies the perturbed "equation"}\\ 0 \in \Phi(z^k)+D\Phi^{t(k)}(z^k)(z-z^k), \quad z^{k+1}:=z.\end{array} \tag{4.6}$$

Here, the parameter t does not appear in $\Phi(z^k)$. So we are dealing with a perturbed Newton method in the sense of Theorem 1 and the function ψ can be used as some regularization of the "Newton matrix".

Clearly, for stabilizing the solution method, we are mostly interested in functions ψ which induce that the new (perturbed) system (1.7) is always solvable (even if the iteration points are far from the solution). Examples will be given below.

Corollary 1 *Suppose that $D\Phi, D\Phi^t$ is defined by one of the derivatives 2...5 of (3.3) and $D\Phi$ fulfills the regularity condition (2.2). Then, for sufficiently small $\|z^1-z^*\|$, the algorithm (4.6) generates a sequence satisfying*

$$\|z^{k+1}-z^*\| \le const\;\|z^k-z^*\|^2.$$

Proof. In all cases, we have $D\Phi^{t(k)}(z^k) \subset D\Phi(z^k)+\|D_z\psi(.,t(k))\|B$ and, due to $f, g_i \in C^{2,1}$ and (4.5), condition (CA) holds in the settings (1.2) and (4.3), for t near the origin, with uniform constants. Because of the choice of $t(k)$, it holds

$$D\Phi^{t(k)}(z^k) \subset D\Phi(z^k)+\|D_z\psi(.,t(k))\|B \subset \; D\Phi(z^k)+C_\psi\|t(k)\|B \subset \; D\Phi(z^k)+C_\psi C\|\Phi(z^k)\|B.$$

Therefore, every matrix $M \in D\Phi^{t(k)}(z^k)$ belongs to $D\Phi(z^k)+A^k$ with some A^k such that $\|A^k\| \le C_\psi C\|\Phi(z^k)\|$, whereafter Theorem 1 ensures the assertion. □

Notice that Theorem 1 can be applied for further estimates.

For understanding the modified Newton steps, we are going to discuss the resulting embedding equation

$$\Phi^t(z)=0 \tag{4.7}$$

which can be also used for obtaining a next iteration point z^{k+1} via

$$0 \;\in\; \Phi^t(z^k)+D\Phi^t(z^k)(z-z^k), \qquad t=t(k). \tag{4.8}$$

This way is the traditional one, denote the resulting solution scheme by *ALG2*. For sufficiently small errors $\|z^k-z^*\|$ the iterates z^{k+1} are close to some path of solutions to $\Phi^t(z)=0$ for $t\to 0$.

However, since both $\Phi(z^k)$ and $D\Phi(z^k)$ have been perturbed in (4.8), the approximation has to be better than in algorithm (4.6), namely

$$\|t(k)\| \le o(\Phi(z^k)), \tag{4.9}$$

in order to ensure superlinear convergence, cf. [10], ch.10.1.1.

To see the necessity of this higher accuracy (in general), it suffices to identify $D\Phi^t(z^k)$ with a fixed regular matrix.

In addition, one may compare the "embedding Newton equation" (4.8) and the condition in (4.6),

$$0 \;\in\; \Phi(z^k)+D\Phi^t(z^k)(z-z^k), \qquad t=t(k),$$

by writing

$$(a,b)^{\mathsf{T}} = \Phi^t(z^k) - \Phi(z^k) = -\psi(z^k,t).$$

This shows that (4.8) means $0 \in [(a,b)^{\mathsf{T}} + \Phi(z^k)] + D\Phi^t(z^k)(z-z^k)$. Thus the latter Newton "equation" (with matrix approximation) is not directly assigned to the original problem but to some of the canonically perturbed problems (1.5).

Next we shall use that (4.7) is always related to a fixed point condition since

$$\Phi^t(z) = 0 \;\Leftrightarrow\; \Phi(z) = \psi(z,t) \;\Leftrightarrow\; z \in H_t(z) := \Phi^{-1}(\psi(z,t)). \tag{4.10}$$

4.2 Existence and behavior of solutions

Theorem 2 (zeros of Φ^t). *Under the general assumptions (4.1) and (4.4), one has:*

(i) There exist positive constants K, ε and δ such that all zeros $z_t = (x_t, y_t)$ of (4.7) with $\|x_t - x^*\| \le \varepsilon$ satisfy

$$\operatorname{dist}((x_t,y_t),(x^*,Y^*)) \le K\|t\| \quad \textit{whenever } \|t\| \le \delta,\ t \in \mathbb{R}^m. \tag{4.11}$$

(ii) Suppose in addition that there is some $\beta > 0$ such that, for all $(a,b) \in B(0,\beta)$, the canonically perturbed problems (1.5) have convex KKT-sets $KKT(a,b)$. Then, for each $\varepsilon > 0$, there is some $\delta > 0$ such that, whenever $\|t\| \le \delta$, some zero z_t of (4.7) with $\|x_t - x^*\| \le \varepsilon$ exists.

(iii) If strong regularity in Robinson's sense [22] is valid at a zero (x^*, y^*) of Φ, then the constants in (i) exist in such a way that related zeros z_t of (4.7) with $\|x_t - x^*\| \le \varepsilon$ uniquely exist and satisfy

$$\|z_{t'} - z_t\| \le K\|t' - t\| \quad \textit{for all } t, t' \in B(0,\delta) \textit{ and } \delta > 0 \textit{ small enough.} \tag{4.12}$$

Proof. The statements (i) and (iii) follow from Corollary 2.9, Corollary 4.4 and Theorem 8.36 in [10], since the map $\psi(\cdot,t)$ is an arbitrary small Lipschitz function in the $C^{0,1}$-norm.

To prove (ii), we simplify the proof in [17], Theorems 2.4, 2.5. By Lemma 1 and Remark 3 there is a compact convex set C such that $\emptyset \ne H(a,b) \subset C \subset \mathbb{R}^{n+m}$ for sufficiently small $\|(a,b)\|$, say for $\|(a,b)\| \le \beta' < \beta$. Moreover, H is closed and convex-valued on $B(0,\beta')$. Define the map

$$H_t^K(z) = H(\psi(z,t)) \subset \mathbb{R}^{n+m}$$

for $z \in C$. If $\|t\| \le \delta$ and δ is small enough, we have $\|\psi(z,t)\| < \beta'$, so H_t^K is again closed and convex-valued and maps C into C. Hence H_t^K has (Kakutani) a fixed point $z^0 \in H_t^K(z^0)$. This means by definition that the assigned Kojima point (x^t, y^t) satisfies (4.10). Using again (4.11) we see that $x_t \to x^*$ as $t \to 0$. □

Remark 4

1. Assertion (ii) of the preceding theorem trivially holds for convex problems, since their KKT-sets are convex.
2. The convexity of the KKT-sets under (ii) can be replaced, due to Remark 3, by supposing that component x for $(x,y) \in \Phi^{-1}(a,b)$ and $\|x - x^*\| \le r$ (with some $r > 0$) is unique.

   For conditions (weaker than strong regularity) which describe this property we refer e.g. to [12] (strong stability in Kojima's sense), [7] (isolated zeros of metrically regular Lipschitz functions) and [11] (strong Lipschitz or Kojima's stability of stationary solutions). However, the given answers are still not complete.
3. Statement (i) has an interesting application for Lipschitz estimates of primal-dual solutions for the standard log-barrier method in the absence of LICQ, see [9].

In §5.1 we shall see that the Lipschitzian inequalities (4.11) and (4.12) may compare solutions of *different methods.* The foregoing theorem says that anyway this can be done in a Lipschitzian manner.

5 Particular parametrizations

The general parametrization includes interesting special cases like

$$\psi_1(z,t) = 0, \quad \psi_{2,i}(z,t) = t_i y_i^+ \tag{5.1}$$

and

$$\psi_1(z,t) = -t_0 x, \quad \psi_{2,i}(z,t) = t_i \textstyle\sum_\nu y_\nu^+ \tag{5.2}$$

or

$$\begin{array}{rcll} \psi_{1,j}(z,t) & = & -t_j \; x_j & j = 1,...,n \\ \psi_{2,i}(z,t) & = & -t_{n+i} \; y_i & i = 1,...,m \end{array} \tag{5.3}$$

It is known that the system (4.7) $\Phi^t(z) = 0$ for (5.1) and (5.2) is closely related to penalty and barrier methods for problem (1.1), see [16, 17] or [10, §11.1]. In the first subsection, we will summarize some related interpretations and transformations from the mentioned literature. In §5.2, we will specify the Newton steps for ALG 1 and ALG 2 discussed above under perturbations (5.1), and this for the cases 2...5 of linearizing the (perturbed) Kojima system.

In the remainder of the paper, we will restrict ourselves to the system (4.7) $\Phi^t(x,y) = 0$ under the perturbation (5.1). To indicate this clearly, we rename Φ^t in this situation by F^t. Hence, we consider for $t \in \mathbb{R}^m$ the zeros of the function

$$\begin{array}{rcccccl} F_1(x,y) & = & Df(x) & + & \sum_{i=1}^m y_i^+ Dg_i(x), & & \\ F_{2,i}^t(x,y) & = & g_i(x) & - & y_i^- & - & t_i y_i^+. \end{array} \tag{5.4}$$

For given t, we know that $F^t(x,y) = M(x)N^t(y)$, where $M(x)$ is the matrix (3.1), and

$$N^t(y) = (1,\; y_1^+, \dots, y_m^+,\; y_1^- + t_1 y_1^+, ..., y_m^- + t_m y_m^+)^\mathsf{T} \in \mathbb{R}^{1+2m}.$$

Following the arguments in §3.2, we then have in the case 2 according to (3.3), (3.4) that the (generalized) derivative at (x,y) has the form

$$DF^t(x,y)(u,v) = \left\{ G(x,y,p,t) \begin{pmatrix} u \\ v \end{pmatrix} \;\middle|\; p \in \mathcal{R}(y) \right\}, \tag{5.5}$$

where again $\mathcal{R}(y) = \{p = (p_1, \dots, p_m) \in [0,1]^m \,|\, p_i = 1 \text{ if } y_i > 0,\; p_i = 0 \text{ if } y_i < 0\}$ and $G(x,y,p,t)$ is the matrix

$$G(x,y,p,t) = \begin{pmatrix} D_x^2 L(x,y^+) & p_1 Dg_1(x) & \dots & p_m Dg_m(x) \\ Dg_1(x)^\mathsf{T} & -(1-p_1+t_1p_1) & & \\ \vdots & & \ddots & \\ Dg_m(x)^\mathsf{T} & & & -(1-p_m+t_mp_m) \end{pmatrix} \tag{5.6}$$

In the cases 3-5 of choosing a generalized derivative according to (3.3), (3.4), the formula (5.5) has to be modified, again by restricting $\mathcal{R}(y)$ to such p satisfying for $y_i = 0$ the special settings $p_i \in \{0,1\}$ (case 3), $p_i = 1$ (case 4) and $p_i = 0$ (case 5), respectively.

5.1 Relations to penalty-barrier functions under perturbation (5.1)

Consider the system $F^t(x,y) = 0$, with F^t in (5.4), and fix the perturbation parameter t.

Quadratic Penalties: Suppose $t_i > 0 \ \forall i$. Then any zero (x,y) of F^t under the perturbation (5.1) has the following property:

The point x is stationary for the well-known penalty function
$P_t(x) = f(x) + \frac{1}{2} \sum_i t_i^{-1}[g_i(x)^+]^2$.

Conversely, if x is stationary for $P_t(x)$, then (x,y) with

$$y_i = t_i^{-1} g_i(x) \text{ for } g_i(x) > 0 \text{ and } y_i = g_i(x) \text{ for } g_i(x) \le 0$$

solves (4.7). Thus applying the penalty method based on $P_t(x)$ for $t_i \downarrow 0$ or solving $F^t = 0$ for the same t, mean exactly the same.

Quadratic and logarithmic barriers: Suppose $t_i < 0 \ \forall i$. Let $z = (x,y)$ solve $F^t(z) = 0$ under the perturbation (5.1) and put $J(y) = \{i \mid y_i > 0\}$. Then (x,y) has the following properties:

$$\begin{array}{l} \text{The point } x \text{ is feasible for (1.1), fulfills } g_i(x) < 0 \ \forall i \in J(y) \\ \text{and is stationary for the function} \\ Q_t(x) = f(x) + \frac{1}{2} \sum_{i \in J(y)} t_i^{-1}[g_i(x)^-]^2. \end{array} \tag{5.7}$$

Conversely, having some x with the properties (5.7), imposed for any index set $J \subset \{1, ..., m\}$, the point (x,y) with

$$y_i = t_i^{-1} g_i(x) \ \ (i \in J) \quad \text{and} \quad y_i = g_i(x) \ \ (i \in \{1, ..., m\} \setminus J)$$

is a zero of F^t. The zeros (x,y) of F^t, for $t_i < 0 \ \forall i$, can be also characterized by logarithmic barriers:

The point x is feasible for (1.1), fulfills $g_i(x) < 0 \ \forall i \in J(y)$
and is stationary for the logarithmic barrier function
$B_{t,y}(x) = f(x) + \ \sum_{i \in J(y)} \ t_i \ (y_i^+)^2 \ \ln(-g_i(x))$.

For $t_i \to -0$, the factors $s_i = t_i \ (y_i^+)^2$ vanish (since y remains bounded due to MFCQ) as required in usual log-barrier settings. However, for inactive constraints $g_i(x^*) < 0$, we obtain $y_i^t < 0$ from convergence of the perturbed Kojima points, hence constraint i is not included in the sum $B_{t,y}(x)$.

Unchanged constraints: If some component of t, say t_1, is zero, then the line $g_1(x) - y_1^-$ in Kojima's function (1.2) remains unchanged. This means that the first constraint $g_1(x) \le 0$ is still explicitly required and the term $y_1^+ Dg_1(x)$ as usually appears in the Lagrange condition.

5.2 Particular Newton realizations under perturbation (5.1)

Let us start with the Newton scheme of ALG2 (4.8) which includes also the standard Newton step when setting $t = 0$. Given any $(x,y,t) \in \mathbb{R}^{n+2m}$, we then look, in the cases 2 ... 5 of the settings (3.3), (3.4) of generalized derivatives DF^t, for a solution (u,v) to the equation

$$0 = F^t(x,y) + G(x,y,p,t)\binom{u}{v}$$

$$= \begin{pmatrix} D_xL(x,y^+) \ + \ D_x^2L(x,y^+)u \ + \ \sum_{i=1}^m p_i v_i Dg_i(x) \\ \vdots \\ g_i(x) \ - \ y_i^- \ - \ t_i y_i^+ \ + \ Dg_i(x)^\mathsf{T} u \ - \ r_i v_i \\ \vdots \end{pmatrix}, \tag{5.8}$$

where $G(x,y,p,t)\binom{u}{v} \in DF^t(x,y)(u,v)$, and p belongs to $\mathcal{R}(y)$ and

$$r_i := 1 - p_i + t_i p_i = 1 - (1-t_i)p_i, \quad i = 1,\dots,m. \tag{5.9}$$

Note that for given (x,y,t) and $p \in \mathcal{R}(y)$ we have

$$\begin{array}{lcccl} y_i < 0 & \Rightarrow & p_i = 0 & , & r_i = 1, \\ y_i > 0 & \Rightarrow & p_i = 1 & , & r_i = t_i, \\ y_i = 0 & \Rightarrow & p_i \in [0,1] & , & r_i = 1 - (1-t_i)p_i. \end{array}$$

This implies that $r_i = 0$ is only possible for $y_i \geq 0$ and $p_i > 0$, namely if

$$(\, y_i = 0 \ \text{ and } \ 1 = (1-t_i)p_i \,) \quad \text{or} \quad (\, y_i > 0 \ \text{ and } \ t_i = 0\,),$$

and that $r_i \neq 0$ yields, by discussing $y_i < 0, \ y_i > 0$ separately,

$$\frac{p_i}{r_i} y_i^- = 0 \quad \text{ and } \quad y_i^+ = \frac{p_i}{r_i}\,(t_i y_i^+ + y_i^-). \tag{5.10}$$

Setting, for $i \in \{1,\dots,m\,\}$,

$$K := \{i \,|\, r_i = 0\}, \quad J := \{i \,|\, r_i \neq 0\,\}, \tag{5.11}$$

then the i-th equation of the second row (5.8) becomes

$$\begin{array}{ll} i \in K: & g_i(x) + Dg_i(x)^\mathsf{T} u \ = \ 0, \\ i \in J: & v_i = \ [g_i(x) + Dg_i(x)^\mathsf{T} u - y_i^- - t_i y_i^+]/r_i. \end{array} \tag{5.12}$$

So the system (5.8) carries over to

$$g_k(x) + Dg_k(x)^\mathsf{T} u = 0 \quad (k \in K) \tag{5.13}$$

along with the linearized Lagrange equation

$$\begin{aligned} 0 \ = \ & D_xL(x,y^+) \ + \ D_x^2L(x,y^+)u \ + \ \textstyle\sum_K p_k v_k Dg_k(x) \\ & + \textstyle\sum_J \ \frac{p_i}{r_i}\ (g_i(x) + Dg_i(x)^\mathsf{T} u - y_i^- - t_i y_i^+) Dg_i(x). \end{aligned} \tag{5.14}$$

Taking (5.10) into account, one has

$$D_xL(x,y^+) - \textstyle\sum_J \ \frac{p_i}{r_i}\ (y_i^- + t_i y_i^+) Dg_i(x) = D_xL(x,y^+) - \sum_J \ y_i^+ Dg_i(x),$$

and, by using the form of $D_xL(x,y^+)$, (5.14) then becomes

$$\begin{aligned} 0 \ = \ & Df(x) + \textstyle\sum_K y_k^+ Dg_k(x) \\ & + D_x^2L(x,y^+)u \ + \ \textstyle\sum_J \frac{p_i}{r_i}\,(g_i(x) + Dg_i(x)^\mathsf{T} u) Dg_i(x) \\ & + \textstyle\sum_K p_k v_k Dg_k(x). \end{aligned} \tag{5.15}$$

The involved term

$$D_x^2 L(x,y^+)u \;+\; \textstyle\sum_J \frac{p_i}{r_i}\,(g_i(x)+Dg_i(x)^\top u)Dg_i(x)$$

is the derivative of the quadratic function

$$Q(u) = \tfrac{1}{2}\,(\;u^\top D_x^2 L(x,y^+)u + \textstyle\sum_J \frac{p_i}{r_i}\,(g_i(x)+Dg_i(x)^\top u)^2\;) \tag{5.16}$$

with a quadratic form that includes deadic products

$$\tfrac{1}{2}\,(\;D_x^2 L(x,y^+) + \textstyle\sum_J \frac{p_i}{r_i}\,Dg_i(x)Dg_i(x)^\top\;).$$

Notice that indices i with $y_i < 0$ (implying $p_i = 0$) do not play any role in this context.

Theorem 3 *Let $(x,y,t) \in \mathbb{R}^{n+2m}$ and $p \in \mathcal{R}(y)$, and let r, K, J and $Q(u)$ be according to (5.9), (5.11) and (5.16), respectively. Then (u,v) solves the Newton equation (5.8) if and only if, with some $\lambda \in \mathbb{R}^{\operatorname{card} K}$, (u,λ) is a KKT point of the quadratic program*

$$\min_u\; Q(u) + c^\top u \qquad s.t. \qquad g_k(x) + Dg_k(x)^\top u = 0 \quad (k \in K), \tag{5.17}$$

where $c = Df(x) + \sum_K y_k^+ Dg_k(x)$. The vector v is then given by the components

$$v_i \;=\; \begin{cases} \lambda_i/p_i\,, & \text{if } i \in K, \\ (g_i(x) + Dg_i(x)^\top u - y_i^- - t_i y_i^+)/r_i\,, & \text{if } i \in J. \end{cases}$$

Moreover, supposing $t_i \neq 0\ \forall i$ and using particularly $p_i \in \{0,1\}$ (like for the cases 3, 4, 5 of (3.3) with derivative (3.4)) then (5.17) is an unconstrained program, i.e., $K = \emptyset$, where the ratios p_i/r_i, $i \in J$, appearing in the definition (5.16) of $Q(u)$ become

$$\frac{p_i}{r_i} = \begin{cases} \frac{1}{t_i} & \text{if } y_i \geq 0 \\ 0 & \text{if } y_i < 0 \end{cases} \quad \text{in case 4}, \qquad \frac{p_i}{r_i} = \begin{cases} \frac{1}{t_i} & \text{if } y_i > 0 \\ 0 & \text{if } y_i \leq 0 \end{cases} \quad \text{in case 5}. \tag{5.18}$$

◇

Proof. We showed by (5.13) and (5.15) that system (5.8) is the KKT system of the problem (5.17), with $\lambda_k = p_k v_k$, $k \in K$, being the Lagrange multipliers of this program. The form of v_i, $i \in J$, follows from (5.12), while $k \in K$ implies $p_k \neq 0$ and hence $v_k = \lambda_k/p_k$. The form of c follows from (5.15).

Moreover, if $p_i \in \{0,1\}$ then (5.9) gives $r_i \in \{t_i, 1\}$, hence $K = \emptyset$, by assumption on t_i. Finally, (5.18) follows via the definition of r_i in (5.9) directly from the settings $p_i = 1$ for $y_i \geq 0$ in case 4 and $p_i = 0$ for $y_i \leq 0$ in case 5. □

Let us finish this section with the Newton scheme of ALG1 (4.6) which requires in the cases 2 ... 5 of the settings (3.3), (3.4) of generalized derivatives DF^t to look for a solution (u,v) of the equation

$$\begin{aligned} 0 \;&=\; F(x,y) + G(x,y,p,t)\tbinom{u}{v} \\ &=\; \begin{pmatrix} D_x L(x,y^+) \;+\; D_x^2 L(x,y^+)u \;+\; \sum_{i=1}^m p_i v_i Dg_i(x) \\ \vdots \\ g_i(x) \;-\; y_i^- \;+\; Dg_i(x)^\top u \;-\; r_i v_i \\ \vdots \end{pmatrix}, \end{aligned} \tag{5.19}$$

where $G(x, y, p, t)\binom{u}{v} \in DF^t(x, y)(u, v)$.

The only difference to the analysis before Theorem 3 consists in the (now disappearing) term $t_i y_i^+$ in the second row. Identifying this term with 0 (particularly in (5.12) and (5.14)) and repeating the above arguments, we obtain due to (5.10),

Theorem 4 *Let $(x, y, t) \in \mathbb{R}^{n+2m}$ and $p \in \mathcal{R}(y)$, and let r, K, J and $Q(u)$ be according to (5.9), (5.11) and (5.16), respectively. Then (u, v) solves the Newton equation (5.19) if and only if, with some $\lambda \in \mathbb{R}^{\mathrm{card}\, K}$, (u, λ) is a KKT point of the quadratic program*

$$\min_u \; Q(u) + c^\mathsf{T} u \qquad s.t. \qquad g_k(x) + Dg_k(x)^\mathsf{T} u = 0 \quad (k \in K), \tag{5.20}$$

where $c = D_x L(x, y^+)$. The vector v is then given by the components

$$v_i \;=\; \begin{cases} \lambda_i / p_i \,, & \text{if } \; i \in K, \\ (g_i(x) + Dg_i(x)^\mathsf{T} u - y_i^-)/r_i \,, & \text{if } \; i \in J. \end{cases}$$

Moreover, supposing $t_i \neq 0 \; \forall i$ and using particularly $p_i \in \{0, 1\}$ (like for the cases 3, 4, 5 of (3.3) with derivative (3.4)) then (5.20) is an unconstrained program, i.e., $K = \emptyset$, and (5.18) similarly holds. $\diamond$

Comparing the form of c in the last theorems, the difference consists in using a reduced Lagrangian in Theorem 3 and the full Lagrangian in Theorem 4.

6 Conclusions

We obtained that perturbed Kojima systems and perturbed Newton steps related to them describe different approaches for finding stationary solutions of nonlinear programs, in particular barrier and penalty methods, sequential quadratic programming and nonsmooth Newton methods, too. The latter can be reduced to the type of derivative being defined for the absolute value at the origin. Here, the simplest settings require the weakest (Newton-) regularity conditions. Transformations between the iterations points of the related methods have been explicitly given. The permitted size of the perturbations depends on whether only the generalized derivative (in the Newton scheme) or the whole function is perturbed. In the first case (ALG1), the perturbations may Lipschitzian decrease, measured by the original error. In the second one (ALG2), superlinearly vanishing parameters must be choosen in order to obtain superlinear local convergence of the whole method. The second situation is less desirable since the "regularization effect" then disappears faster.

Acknowledgement. The authors wish to thank Stephan Bütikofer (University of Zurich) and an anonymous referee for their critical comments.

References

[1] J.-P. Aubin and I. Ekeland. *Applied Nonlinear Analysis.* Wiley, New York, 1984.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

[3] F.H. Clarke. On the inverse function theorem. *Pacific Journal of Mathematics*, 64: 97–102, 1976.

[4] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983.

[5] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementary Problems, Vol I and Vol II.* Springer, New York, 2003.

[6] A.V. Fiacco and G.P. McCormick. *Nonlinear programming: Sequential unconstrained minimization techniques* Wiley, New York, 1968.

[7] P. Fusek. Isolated zeros of Lipschitzian metrically regular R^n functions. *Optimization* 49: 425–446, 2001.

[8] A. Griewank. The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces. *SIAM Journal on Numerical Analysis*, 24: 684–705, 1987.

[9] C. Grossmann, D. Klatte, and B. Kummer. Convergence of primal-dual solutions for the nonconvex log-barrier method without LICQ. *Kybernetika*, 40:571–584, 2004.

[10] D. Klatte and B. Kummer. *Nonsmooth equations in optimization - regularity, calculus, methods and applications*. Kluwer, Dordrecht, 2002.

[11] D. Klatte and B. Kummer. *Strong Lipschitz Stability of Stationary Solutions for Nonlinear Programs and Variational Inequalities*. *SIAM Journal on Optimization*, 16: 96–119, 2005.

[12] 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.

[13] M. Kojima and S. Shindoh. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *Journal of the Operational Research Society of Japan*, 29:352–372, 1987.

[14] B. Kummer. Newton's method for non-differentiable functions. In J. Guddat et al., editor, *Advances in Math. Optimization*, pages 114–125. Akademie Verlag Berlin, (Ser. Math. Res. 45), Berlin, 1988.

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

[16] B. Kummer. On solvability and regularity of a parametrized version of optimality conditions. *ZOR Mathematical Methods of OR*, 41:215–230, 1995.

[17] B. Kummer. Parametrizations of Kojima's system and relations to penalty and barrier functions. *Mathematical Programming, Series B*, 76:579–592, 1997.

[18] B. Kummer. Generalized Newton and NCP-methods: Convergence, regularity, actions. *Discussiones Mathematicae - Differential Inclusions*, 20:209–244, 2000.

[19] R. Miffin. Semismooth and semiconvex functions in constrained optimization. SIAM J. Control and Optim. 15: 957-972, 1977.

[20] F. Nožička, J. Guddat, H. Hollatz, and B. Bank. *Theorie der linearen parametrischen Optimierung*. Akademie–Verlag, Berlin, 1974.

[21] L. Qi, J. Sun. A nonsmooth version of Newton's method. Math. Programming 58: 353-367, 1993.

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

[23] S.M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis*, 2:291–305, 1994.

[24] L. Thibault. Subdifferentials of compactly Lipschitz vector-valued functions. *Annali di Matematica Pura ed Applicata*, 4:157–192, 1980.

[25] L. Thibault. On generalized differentials and subdifferentials of Lipschitz vector-valued functions. *Nonlinear Analysis: Theory, Methods & Applications*, 6:1037–1053, 1982.