Math. Program., Ser. B (2009) 117:355–386
DOI 10.1007/s10107-007-0173-x

FULL LENGTH PAPER

An accelerated Newton method for equations with semismooth Jacobians and nonlinear complementarity problems

Christina Oberlin · Stephen J. Wright

Received: 28 April 2006 / Accepted: 18 April 2007 / Published online: 10 August 2007

Abstract We discuss local convergence of Newton's method to a singular solution x^* of the nonlinear equations $F(x) = 0$, for $F : \mathbb{R}^n \to \mathbb{R}^n$. It is shown that an existing proof of Griewank, concerning linear convergence to a singular solution x^* from a starlike domain around x^* for F twice Lipschitz continuously differentiable and x^* satisfying a particular regularity condition, can be adapted to the case in which F' is only strongly semismooth at the solution. Further, Newton's method can be accelerated to produce fast linear convergence to a singular solution by overrelaxing every second Newton step. These results are applied to a nonlinear-equations reformulation of the nonlinear complementarity problem (NCP) whose derivative is strongly semismooth when the function f arising in the NCP is sufficiently smooth. Conditions on f are derived that ensure that the appropriate regularity conditions are satisfied for the nonlinear-equations reformulation of the NCP at x^*.

Keywords Nonlinear equations · Semismooth functions · Newton's method · Nonlinear complementarity problems

We dedicate this paper to Steve Robinson on the occasion of his 65th birthday, in recognition of his remarkable scholarly accomplishments and in appreciation for his guidance and his collegiality, grace, and kindness.

Research supported by NSF Grants SCI-0330538, DMS-0427689, CCF-0430504, CTS-0456694, CNS-0540147, DOE grant DE-FG02-04ER25627, and an NSF Graduate Research Fellowship.

C. Oberlin (✉) · S. J. Wright
Computer Sciences Department, University of Wisconsin, 1210 W. Dayton Street, Madison, WI 53706, USA
e-mail: coberlin@cs.wisc.edu

S. J. Wright
e-mail: swright@cs.wisc.edu

Mathematics Subject Classification (2000) 65H10 · 90C33

1 Introduction

Consider a mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, and let $x^* \in \mathbb{R}^n$ be a solution to $F(x) = 0$. We consider the local convergence of Newton's method when the solution x^* is *singular* (that is, $\ker F'(x^*) \neq \{0\}$) and when F is once but possibly not twice differentiable. We also consider an accelerated variant of Newton's method that achieves a fast linear convergence rate under these conditions. Our technique can be applied to a nonlinear-equations reformulation of nonlinear complementarity problems (NCP) defined by

$$0 \leq f(x), \quad x \geq 0, \quad x^T f(x) = 0, \qquad \text{NCP}(f) \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$. At degenerate solutions of the NCP (solutions x^* at which $x_i^* = f_i(x^*) = 0$ for some $i = 1, 2, \ldots, n$), the nonlinear-equations reformulation considered in this paper is not twice differentiable at x^*, and weaker smoothness assumptions are required. Our results show that (i) Newton's method applied to the nonlinear-equations reformulation of the NCP converges linearly inside a "starlike domain" centered at a singular solution x^*; (ii) a simple technique can be applied to achieve a faster linear rate. The simplicity of our approach contrasts with other nonlinear-equations-based approaches to solving (1), which are either nonsmooth (and hence require nonsmooth Newton techniques whose implementations are more complex; see for example Josephy [14], Facchinei and Pan [6, p. 663–674] or else require classification of the indices $i = 1, 2, \ldots, n$ into those for which $x_i^* = 0$, those for which $f_i(x^*) = 0$, or both.

Around 1980, several authors, including Reddien [20], Decker and Kelley [3], and Griewank [8], proved linear convergence for Newton's method to a singular solution x^* of F from special regions near x^*, provided that F is twice Lipschitz continuously differentiable and a certain 2-regularity condition holds at x^*. (The "2" emphasizes the role of the second derivative of F in this regularity condition.) In the first part of this work, we show that Griewank's analysis, which gives linear convergence from a starlike domain of x^* (defined below), can be extended to the case in which F' is strongly semismooth at x^*; see Sect. 4. In Sect. 5, we consider a standard acceleration scheme for Newton's method, which "overrelaxes" every second Newton step. By assuming that F' is at least strongly semismooth at x^* and that a 2-regularity condition holds, we show that this technique yields arbitrarily fast linear convergence from a partial neighborhood of x^*.

In the second part of this work, beginning in Sect. 6, we consider a nonlinear-equations reformulation of the NCP and interpret the regularity conditions for this reformulation as conditions on the NCP. We show that they reduce to previously known NCP regularity conditions in certain cases. We conclude in Sect. 7 by presenting computational results for some simple NCPs, including a number of degenerate examples.

We start with definitions and terminology (Sect. 2) and a discussion of prior relevant work (Sect. 3).

2 Definitions and properties

For $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$, we follow convention in writing the derivative G' as a map from $\Omega \to \mathbb{R}^{p\times n}$ when $p > 1$ and a map from Ω to $\mathbb{R}^n$ (equivalently $\mathbb{R}^{n\times 1}$) when $p = 1$. The Euclidean norm is denoted by $\|\cdot\|$, and the unit sphere is $\mathcal{S} = \{t \mid \|t\| = 1\}$.

For any subspace X of $\mathbb{R}^n$, dim X denotes the dimension of X. The kernel of a linear operator M is denoted ker M, the image or range of the operator is denoted range M. rank M denotes the rank of the matrix M, which is the dimension of range M.

A *starlike domain* with respect to $x^* \in \mathbb{R}^n$ is an open set $\mathcal{A}$ with the property that $x \in \mathcal{A} \Rightarrow \lambda x + (1-\lambda)x^* \in \mathcal{A}$ for all $\lambda \in (0, 1)$. A vector $t \in \mathcal{S}$ is an *excluded direction* for $\mathcal{A}$ if $x^* + \lambda t \notin \mathcal{A}$ for all $\lambda > 0$.

We now list various definitions relating to the smoothness of a function.

Definition 1 Directionally differentiable. Let $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$, with Ω open, $x \in \Omega$, and $d \in \mathbb{R}^n$. If the limit

$$\lim_{t\downarrow 0} \frac{G(x+td) - G(x)}{t} \tag{2}$$

exists in $\mathbb{R}^p$, G has a *directional derivative* at x along d and we denote this limit by $G'(x; d)$. G is *directionally differentiable* at x if $G'(x; d)$ exists for every d in a neighborhood of the origin.

Definition 2 B-differentiable. ([6, Definition 3.1.2]) $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$, with Ω open, is *B(ouligand)-differentiable* at $x \in \Omega$ if G is Lipschitz continuous in a neighborhood of x and directionally differentiable at x.

Definition 3 Strongly semismooth. ([6, Definition 7.4.2]) Let $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$, with Ω open, be locally Lipschitz continuous on Ω. G is *strongly semismooth* at $\bar{x} \in \Omega$ if G is directionally differentiable near $\bar{x}$ and

$$\limsup_{\bar{x}\neq x\to\bar{x}} \frac{\|G'(x; x-\bar{x}) - G'(\bar{x}; x-\bar{x})\|}{\|x-\bar{x}\|^2} < \infty.$$

Further, G is strongly semismooth on Ω if G is strongly semismooth at every $\bar{x} \in \Omega$.

If G is (strongly) semismooth at $\bar{x}$, then it is B-differentiable at $\bar{x}$. Further, if G is B-differentiable at $\bar{x}$, then $G'(\bar{x}; \cdot)$ is Lipschitz continuous [19]. Hence, for $F' : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ strongly semismooth at x^*, there is some L_{x^*} such that

$$\|(F')'(x^*; h_1) - (F')'(x^*; h_2)\| \le L_{x^*}\|h_1 - h_2\|. \tag{3}$$

If F' is strongly semismooth at x^* and $\|x - x^*\|$ is sufficiently small, we have the following crucial estimate from equation (7.4.5) of [6]:

$$F'(x) = F'(x^*) + (F')'(x^*; x - x^*) + O(\|x - x^*\|^2). \tag{4}$$

(We use $p = n^2$ in order to apply Definition 3 to F'.)

Lastly, we define 2-regularity and its variants. For $F : \mathbb{R}^n \to \mathbb{R}^n$, suppose x^* is a singular solution of $F(x) = 0$ and F' is strongly semismooth at x^*. We define $N := \ker F'(x^*)$. Let $N_\perp$ denote the orthogonal complement of N, such that $N \oplus N_\perp = \mathbb{R}^n$, and let $N_* := \ker F'(x^*)^T$ with orthogonal complement $N_{*\perp}$. We denote by P_N, $P_{N_\perp}$, P_{N_*}, and $P_{N_{*\perp}}$ the orthogonal projections onto N, $N_\perp$, N_*, and $N_{*\perp}$ respectively, while $(\cdot)|_N$ denotes the restriction map to N. Let $m := \dim N > 0$.

We say that F satisfies *2-regularity* for some $d \in \mathbb{R}^n$ at a solution x^* if

$$(P_{N_*} F')'(x^*; d)|_N \text{ is nonsingular.} \tag{5}$$

The 2-regularity conditions of Reddien [20], Decker and Kelley [3], and Griewank [8] require (5) to hold for certain $d \in N$. In fact, this property first appeared in the literature as nonsingularity of $(P_{N_*} F''(x^*)d)|_N$; the form in (5) was introduced by Izmailov and Solodov [11]. By applying P_{N_*} to F' before taking the directional derivative, the theory of 2-regularity may be applied to problems for which $P_{N_*} F'$ is directionally differentiable but F' is not (see [13]).

Decker and Kelley [3] and Reddien [20] use the following definition of 2-regularity, which we call $2^\forall$-regularity.

Definition 4 $2^\forall$-regularity. $2^\forall$*-regularity* holds for F at x^* if (5) holds for every $d \in N \setminus \{0\}$.

For F twice differentiable at x^*, $2^\forall$-regularity implies (geometric) isolation of the solution x^* [3,20] and limits the dimension of N to at most 2 [4].

Next, we define a weaker 2-regularity that can hold regardless of the dimension of N or whether x^* is isolated.

Definition 5 2^{ae}-regularity. 2^{ae}*-regularity* holds for F at x^* if (5) holds for almost every $d \in N$.

Weaker still is the condition we call 2^1-regularity.

Definition 6 2^1-regularity. 2^1*-regularity* holds for F at x^* if (5) holds for some $d \in N$.

For the case in which F is twice Lipschitz continuously differentiable, Griewank shows that 2^1-regularity and 2^{ae}-regularity are actually equivalent [8, p. 110]. This property fails to hold under the weaker smoothness conditions of this work. For example, the smooth nonlinear equations reformulation (7) of the nonlinear complementarity problems quad2 and affknot1 (defined in Appendix A) are 2^1-regular but not 2^{ae}-regular at their solutions.

Izmailov and Solodov [11, Theorem 5(a)] introduce a regularity condition and prove that it implies isolation of the solution, provided that $P_{N_*} F'(x^*)$ is B-differentiable . The following form of this condition, which we call 2^T-regularity, is specific to our case $F : \mathbb{R}^n \to \mathbb{R}^n$ and is due to Daryina, Izmailov, and Solodov [1, Def. 2.1]. Consider the set

$$T_2 := \{d \in N \mid (P_{N_*} F')'(x^*; d)d = 0\}. \tag{6}$$

Definition 7 2^T-regularity [1, Def. 2.1]. 2^T *-regularity* holds for F at x^* if $T_2 = \{0\}$.

As can be seen from Table 1 in Sect. 7, neither 2^T-regularity nor 2^{ae}-regularity implies the other condition. By definition, $2^\forall$-regularity implies the other three regularity conditions. Therefore, since 2^T-regularity implies isolation of the solution under our smoothness conditions, so must $2^\forall$-regularity. If dim $N = 1$, then 2^T-regularity is equivalent to $2^\forall$-regularity (which is trivially equivalent to 2^{ae}-regularity in this case).

3 Prior work

In this section, we summarize prior work relevant to this paper.

2-regularity conditions. 2-regularity has been applied to a variety of uses including error bounds, implicit function theorems, and optimality conditions [11,13]. We focus here on the use of 2-regularity conditions to prove convergence of Newton-like methods to singular solutions.

The 2^1-regularity condition (Definition 6) was used by Reddien [21] and Griewank and Osborne [10]. The proofs therein show convergence of Newton's method (at a linear rate of $1/2$) only for starting points x_0 such that $x_0 - x^*$ lies approximately along the particular direction d for which the nonsingularity condition (5) holds. The more stringent $2^\forall$-regularity condition (Definition 4) was used by Decker and Kelley [3] to prove linear convergence of Newton's method from starting points in a particular truncated cone around N. The convergence analysis given for $2^\forall$-regularity [2,3,20] is much simpler than the analysis presented by Griewank [8], and in the current paper.

Griewank [8] proves convergence of Newton's method from all starting points in a starlike domain with respect to x^*. If 2^1-regularity holds at x^* and F is twice Lipschitz continuously differentiable at x^*, then the starlike domain is "dense" near x^* in the sense that the set of excluded directions has measure zero—a much more general set than the cones around N analyzed prior to that time.

Acceleration techniques. When iterates $\{x_k\}$ generated by a Newton-like method converge to a singular solution, the error $x_k - x^*$ lies predominantly in the null space N of $F'(x^*)$. Acceleration schemes typically attempt to stay within a cone around N while lengthening ("overrelaxing") some or all of the Newton steps.

We discuss several of the techniques proposed in the early 1980s. All require starting points whose error lies in a cone around N, and all assume three times differentiability of F. Decker and Kelley [4] prove superlinear convergence for a scheme in which every second Newton step is essentially doubled in length along the subspace N. Their technique requires $2^\forall$-regularity at x^*, an estimate of N, and a nonsingularity condition over N on the third derivative of F at x^*. Decker, Keller, and Kelley [2] prove superlinear convergence when every third step is overrelaxed, provided that 2^1-regularity holds at x^* and the third derivative of F at x^* satisfies a nonsingularity condition on N. Kelley and Suresh [16] require 2^1-regularity at x^* and the third derivative of F at x^* bounded over the truncated cone about N. Overrelaxing every other step by a factor approaching 2 results in superlinear convergence.

By contrast, the acceleration technique given in Sect. 5 of our paper does not require the starting point x_0 to be in a cone about N, and requires only strong semismoothness of F' at x^*. On the other hand, we obtain only fast linear convergence. We believe, however, that our analysis can be extended to a superlinear scheme like that of Kelley and Suresh [16].

Smooth nonlinear-equations reformulation of the NCP. In the latter part of this paper, we discuss a nonlinear-equations reformulation of the NCP Ψ based on the function $\psi_s(a, b) := 2ab - (\min(0, a + b))^2$, which has the property that $\psi_s(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$, and $ab = 0$. The function $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\Psi_i(x) := 2x_i f_i(x) - [\min(0, x_i + f_i(x))]^2, \quad i = 1, 2, \ldots, n. \tag{7}$$

This reformulation is apparently due to Evtushenko and Purtov [5] and was studied further by Kanzow [15]. The first derivative Ψ' is strongly semismooth at a solution x^* if f' is strongly semismooth at x^*. If $x_i^* = f_i(x^*) = 0$ for some i, x^* is a singular root of Ψ, which fails to be twice differentiable.

Recently, Izmailov and Solodov [11–13] and Daryina, Izmailov, and Solodov [1] have investigated the properties of the mapping Ψ and designed algorithms around it. (Some of their investigations, like ours, have taken place in the more general setting of a mapping F for which F' has semismoothness properties.) Izmailov and Solodov [11,13] show that an error bound for NCPs holds whenever 2^T-regularity holds. Using this error bound to classify the indices $i = 1, 2, \ldots, n$, Daryina, Izmailov, and Solodov [1] present an active-set Gauss-Newton-type method for NCPs. They prove superlinear convergence to singular points which satisfy 2^T-regularity as well as another condition known as weak regularity, which requires full rank of a certain submatrix of $f'(x^*)$. In [12], Izmailov and Solodov augment the reformulation $\Psi(x) = 0$ by adding a nonsmooth function containing second-order information. They apply the generalized Newton's method to the resulting function and prove superlinear convergence under 2^T-regularity and another condition called quasi-regularity, discussed further in Subsect. 6.3 below.

In contrast to the algorithms of [1] and [12], our approach has fast linear convergence rather than superlinear convergence. Our regularity conditions are comparable and may be weaker in some cases. (For example, the problem munson4 in Appendix A satisfies both 2^T-regularity and 2^{ae}-regularity but not weak regularity.) We believe that our algorithm has the advantage of simplicity. Near the solution, it modifies Newton's method only by incorporating a simple check to detect linear convergence and possibly overrelaxing every second step. There is no need to classify the constraints, add "bordering" terms, or switch to a different step computation strategy in the final iterations.

4 Convergence of the Newton step to a singularity

Griewank [8] extended the work of others [3,20] to prove local convergence of Newton's method from a starlike domain $\mathcal{R}$ of a singular solution x^* of $F(x) = 0$.

Specialized to the case of $k = 1$ (Griewank's notation), he assumes that $F''(x)$ is Lipschitz continuous near x^* and x^* is a 2^1-regular solution. Griewank's convergence analysis shows that the first Newton step takes the initial point x_0 from the original starlike domain $\mathcal{R}$ into a simpler starlike domain $\mathcal{W}_s$, a wedge around a vector s in the null space N. The domain $\mathcal{W}_s$ is similar to the domains of convergence found in earlier works [3,20]. Linear convergence is then proved inside $\mathcal{W}_s$. For F twice continuously differentiable, the convergence domain $\mathcal{R}$ is much larger than $\mathcal{W}_s$. In fact, the set of directions excluded from $\mathcal{R}$ has zero measure.

We weaken the smoothness assumption of Griewank in replacing the second derivative of F in (5) by a directional derivative of F'. Our assumptions follow:

Assumption 1 For $F : \mathbb{R}^n \to \mathbb{R}^n$, x^* is a singular, 2^1-regular solution of $F(x) = 0$ and F' is strongly semismooth at x^*.

We show that Griewank's convergence results hold under this assumption.

Theorem 1 *Suppose Assumption 1 holds. There exists a starlike domain $\mathcal{R}$ about x^* such that, if Newton's method for $F(x)$ is initialized at any $x_0 \in \mathcal{R}$, the iterates converge linearly to x^* with rate $1/2$. If the problem is converted to standard form (8) and $x_0 = \rho_0 t_0$, where $\rho_0 = \|x_0\| > 0$ and $t_0 \in \mathcal{S}$, then the iterates converge inside a cone with axis $g(t_0)/\|g(t_0)\|$, for g defined in (30).*

Only a few modifications to Griewank's proof [8] are necessary. We use the properties (3) and (4) to show that F is smooth enough for the steps in the proof to hold. Finally, we make an insignificant change to a constant required by the proof due to a loss of symmetry in $\mathcal{R}$. (Symmetry is lost in moving from derivatives to directional derivatives because directional derivatives are positively but not negatively homogeneous.) The proof in [8] also considers regularities larger than 2, for which higher derivatives are required. We restrict our discussion to 2-regularity because we are interested in the application to a nonlinear-equations reformulation of NCP, for which such higher derivatives are unavailable.

In the remainder of this section, we develop some preliminaries, discuss domains of invertibility of the Jacobian and convergence of the Newton iterates, analyze the form of a Newton step, and finally sketch the proof of Theorem 1. The proof is presented in full in the extended technical report [18], where its points of departure from Griewank's proof are highlighted.

4.1 Preliminaries

For simplicity of notation, we start by standardizing the problem. The Newton iteration is invariant with respect to nonsingular linear transformations of F and nonsingular affine transformations of the variables x. As a result, we can assume that

$$x^* = 0, \quad F'(x^*) = F'(0) = (I - P_{N_*}), \text{ and } N_* = \mathbb{R}^m \times \{0\}^{n-m}. \tag{8}$$

(We revoke assumption (8) in our discussion of an equation reformulation of the NCP in Sects. 6 and 7.)

For $x \in \mathbb{R}^n \setminus \{0\}$, we write $x = x^* + \rho t = \rho t$, where $\rho = \|x\|$ is the 2-norm distance to the solution and $t = x/\rho$ is a direction in the unit sphere $\mathcal{S}$. From the third assumption in (8), we have

$$P_{N_*} = \begin{bmatrix} I_{m\times m} & 0_{m\times n-m} \\ 0_{n-m\times m} & 0_{n-m\times n-m} \end{bmatrix},$$

where I represents the identity matrix and 0 the zero matrix, with subscripts indicating their dimensions. By substituting in the second assumption of (8), we obtain

$$F'(0) = \begin{bmatrix} 0_{m\times m} & 0_{m\times n-m} \\ 0_{n-m\times m} & I_{n-m\times n-m} \end{bmatrix}. \tag{9}$$

Since $F'(0)$ is symmetric, the null space N is identical to N_*.

Using (8), we partition $F'(x)$ as follows:

$$F'(x) = \begin{bmatrix} P_{N_*}F'(x)|_N & P_{N_*}F'(x)|_{N_\perp} \\ P_{N_{*\perp}}F'(x)|_N & P_{N_{*\perp}}F'(x)|_{N_\perp} \end{bmatrix} =: \begin{bmatrix} B(x) & C(x) \\ D(x) & E(x) \end{bmatrix}.$$

In conformity with the partitioning in (9), the submatrices B, C, D, and E have dimensions $m \times m, m \times n-m, n-m \times m$, and $n-m \times n-m$, respectively. Using $x^* = 0$, we define

$$\bar{B}(x) = \bar{B}(x - x^*) = (P_{N_*}F')'(x^*; x - x^*)|_N = (P_{N_*}F')'(0; x)|_N, \tag{10a}$$

$$\bar{C}(x) = \bar{C}(x - x^*) = (P_{N_*}F')'(x^*; x - x^*)|_{N_\perp} = (P_{N_*}F')'(0; x)|_{N_\perp}. \tag{10b}$$

From $x = \rho t$, the expansion (4) with $x^* = 0$ yields

$$\begin{aligned} B(x) &= \bar{B}(x) + O(\rho^2) = \rho\bar{B}(t) + O(\rho^2), \\ C(x) &= \bar{C}(x) + O(\rho^2) = \rho\bar{C}(t) + O(\rho^2), \\ D(x) &= O(\rho), \quad \text{and} \quad E(x) = I + O(\rho). \end{aligned} \tag{11}$$

The constants that bound the $O(\cdot)$ terms in these expressions can be chosen independently of t, by compactness of $\mathcal{S}$. This is the first difference between our analysis and Griewank's analysis: We use (4) to arrive at (11), while he uses Taylor's theorem.

For some $r_b > 0$, E is invertible for all $\rho < r_b$ and all $t \in \mathcal{S}$, with $E^{-1}(x) = I + O(\rho)$. Invertibility of $F'(x)$ is equivalent to invertibility of the Schur complement of $E(x)$ in $F'(x)$, which we denote by $G(x)$ and define by

$$G(x) := B(x) - C(x)E(x)^{-1}D(x).$$

This claim follows from the determinant formula $\det(F'(x)) = \det(G(x))\det(E(x))$. By reducing r_b if necessary to apply (11), we have

$$G(x) = B(x) + O(\rho^2) = \rho\bar{B}(t) + O(\rho^2). \tag{12}$$

Hence,

$$\det(F'(x)) = \rho^m \det\bar{B}(t) + O\left(\rho^{m+1}\right).$$

As in the proof of [8, Lemma 3.1(iii)], we note that all but the smallest m singular values of $F'(x)$ are close to 1 in a neighborhood of x^*. Letting $\nu(s)$ denote the smallest singular value of $F'(s)$, we have that

$$\nu(\rho t) = O((\det F'(\rho t))^{1/m}) = \begin{cases} O(\rho), & \text{if } \bar{B}(t) \text{ is nonsingular,} \\ o(\rho), & \text{if } \bar{B}(t) \text{ is singular.} \end{cases} \tag{13}$$

For later use, we define γ to be the smallest positive constant such that

$$\|G(x) - \rho\bar{B}(t)\| \leq \gamma\rho^2, \quad \text{for all } x = \rho t, \text{ all } t \in \mathcal{S}, \text{ and all } \rho < r_b.$$

Following Griewank [8], we define the function $\sigma(t)$ to be the minimum of 1 and the smallest (in magnitude) singular value of $\bar{B}(t)$, that is,

$$\sigma(t) := \begin{cases} 0, & \text{if } \bar{B}(t) \text{ is singular} \\ \min(1, \|\bar{B}^{-1}(t)\|^{-1}), & \text{otherwise.} \end{cases} \tag{14}$$

The individual singular values of a matrix vary continuously with respect to perturbations of the matrix [7, Theorem 8.6.4]. By (3), $\bar{B}(t)$ is Lipschitz continuous in t, so that $\sigma(t)$ is continuous in t. This is the second difference between our analysis and Griewank's analysis: we require (3) to prove continuity of the singular values of $\bar{B}(t)$, while he uses the fact that $\bar{B}(t)$ is linear in t, which holds under his smoothness assumptions.

Let

$$\Pi_0(d) := \det\bar{B}(d), \qquad \text{for } d \in \mathbb{R}^n. \tag{15}$$

In contrast to the smooth case considered by Griewank, $\Pi_0(d)$ is not a homogeneous polynomial in d, but rather a positively homogeneous, piecewise-smooth function. Hence, 2^1-regularity does not imply 2^{ae}-regularity. Since the determinant is the product of singular values, we can use the same reasoning as for $\sigma(t)$ to deduce that $\Pi_0(t)$ is continuous in t for $t \in \mathcal{S}$.

4.2 Domains of invertibility and convergence

In this section we define the domains $\mathcal{W}_s$ and $\mathcal{R}$. These definitions depend on several functions that we now introduce. If we define $\min(\emptyset) = \pi$, the angle

$$\phi(s) := \frac{1}{4}\min\{\cos^{-1}(t^T s) \mid t \in \mathcal{S} \cap {\Pi_0}^{-1}(0)\}, \quad \text{for } s \in N \cap \mathcal{S} \tag{16}$$

is a well defined, nonnegative continuous function, bounded above by $\frac{\pi}{4}$. For the smooth case considered by Griewank, if $t \in {\Pi_0}^{-1}(0)$, then $-t \in {\Pi_0}^{-1}(0)$ and the maximum angle if ${\Pi_0}^{-1}(0) \neq \emptyset$ is $\frac{\pi}{2}$. This is no longer true in our case; the corresponding maximum angle is π. Hence, we have defined $\min(\emptyset) = \pi$ (instead of Griewank's definition $\min(\emptyset) = \frac{\pi}{2}$) and the coefficient of $\phi(s)$ is $\frac{1}{4}$ instead of $\frac{1}{2}$. (This is the third and final difference between our analysis and Griewank's.) Now, $\phi^{-1}(0) = N \cap \mathcal{S} \cap {\Pi_0}^{-1}(0)$ because the set $\{s \in \mathcal{S} \mid \Pi_0(s) \neq 0\}$ is open in $\mathcal{S}$ since $\Pi_0(\cdot)$ is continuous on $\mathcal{S}$, by (3).

In [8, Lemma 3.1], Griewank defines the auxiliary starlike domain of invertibility $\bar{\mathcal{R}}$,

$$\bar{\mathcal{R}} := \{x = \rho t \mid t \in \mathcal{S},\ 0 < \rho < \bar{r}(t)\}, \tag{17}$$

where

$$\bar{r}(t) := \min\left\{r_b, \frac{1}{2}\gamma^{-1}\sigma(t)\right\}. \tag{18}$$

The excluded directions of $\bar{R}$, $t \in \mathcal{S}$ for which $\sigma(t) = 0$, are the directions along which the smallest singular value of the determinant of $F'(\rho t)$ is $o(\rho)$, by (13) and (14). This set of directions may have measure that is positive but less than one in $\mathcal{S}$. This is the case for the smooth nonlinear equations reformulation (7) of the nonlinear complementarity problem quad2 (defined in Appendix A). For this problem, $\sigma(t) \neq 0$ for almost every $t = (t^1, t^2)^T \in \mathcal{S}$ with $t^1 < 0$ and $t^2 \neq 0$, while $\sigma(t) = 0$ for any $t \in \mathcal{S}$ with $t^1 > 0$.

As in [8, Lemma 5.1], we define

$$\hat{r}(s) := \min\{\bar{r}(t) \mid t \in \mathcal{S},\ \cos^{-1}(t^T s) \leq \phi(s)\}, \quad \text{for } s \in N \cap \mathcal{S}, \tag{19}$$

$$\hat{\sigma}(s) := \min\{\sigma(t) \mid t \in \mathcal{S},\ \cos^{-1}(t^T s) \leq \phi(s)\}, \quad \text{for } s \in N \cap \mathcal{S}. \tag{20}$$

These minima exist and both are nonnegative and continuous on $\mathcal{S} \cap N$ with $\hat{\sigma}^{-1}(0) = \hat{r}^{-1}(0) = \phi^{-1}(0)$. Since $\sigma(t) \leq 1$ by definition, we have $\hat{\sigma}(s) \leq 1$ for $s \in N \cap \mathcal{S}$.

Let c be the positive constant defined by

$$c := \max\{\|\bar{C}(t)\| + \sigma(t) \mid t \in \mathcal{S}\}. \tag{21}$$

In the following, we use the abbreviation

$$q(s) := \frac{1}{4}\sin\phi(s) \leq \frac{1}{4}, \text{ for } s \in N \cap \mathcal{S}. \tag{22}$$

We define the angle $\hat{\phi}(s)$, for which $0 \leq \hat{\phi}(s) \leq \pi/2$, by the equality

$$\sin\hat{\phi}(s) := \min\left\{\frac{q(s)}{c/\hat{\sigma}(s) + 1 - q(s)}, \frac{2\delta\hat{r}(s)}{(1 - q(s))\hat{\sigma}^2(s)}\right\}, \text{ for } s \in N \cap \mathcal{S}, \tag{23}$$

where δ is a problem-dependent, positive number specified in (39). We define

$$\hat{\rho}(s) := \frac{(1 - q(s))\hat{\sigma}^2(s)}{2\delta}\sin\hat{\phi}(s), \text{ for } s \in N \cap \mathcal{S}. \tag{24}$$

Both $\hat{\phi}$ and $\hat{\rho}$ are nonnegative and continuous on $N \cap \mathcal{S}$ with

$$\hat{\phi}^{-1}(0) = \hat{\rho}^{-1}(0) = \phi^{-1}(0) = \Pi_0^{-1}(0) \cap N \cap \mathcal{S}. \tag{25}$$

Now we can define the starlike domains $\mathcal{W}_s$ and $\mathcal{I}_s$ as follows:

$$\mathcal{W}_s := \{x = \rho t \mid t \in \mathcal{S},\ \cos^{-1}(t^T s) < \hat{\phi}(s), 0 < \rho < \hat{\rho}(s)\}, \tag{26}$$

$$\mathcal{I}_s := \{x = \rho t \mid t \in \mathcal{S}, \cos^{-1}(t^T s) < \phi(s), 0 < \rho < \hat{\rho}(s)\}. \tag{27}$$

By the first inequality in (23), $\sin \hat{\phi}(s) \leq \sin \phi(s)$. Since both $\hat{\phi}(s)$ and $\phi(s)$ are acute angles, we have $\hat{\phi}(s) \leq \phi(s)$ and thus $\mathcal{W}_s \subseteq \mathcal{I}_s$. For $s \in \mathcal{S} \cap N$, $\mathcal{W}_s = \emptyset$ if and only if $\Pi_0(s) = 0$. The second implicit inequality in the definition of $\sin \hat{\phi}(s)$, ensures that $\hat{\rho}(s)$ satisfies

$$\hat{\rho}(s) \leq \hat{r}(s) \leq \bar{r}(t) \leq r_b, \quad \text{for all } t \in \mathcal{S} \text{ with } \cos^{-1} t^T s \leq \phi(s). \tag{28}$$

It follows that

$$\mathcal{I}_s \subset \bar{\mathcal{R}}, \quad \text{for all } \ s \in \mathcal{S} \cap N \setminus \Pi_0^{-1}(0). \tag{29}$$

(The justification given in [8] that $\hat{r}(s) \leq \bar{r}(s)$ is insufficient.)

Consider the positively homogeneous vector function $g : (\mathbb{R}^n \setminus \Pi_0^{-1}(0)) \to N \subseteq \mathbb{R}^n$,

$$g(x) = \rho g(t) = \begin{bmatrix} I & \bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x. \tag{30}$$

It is shown in (40) that the Newton iteration from a point x near $x^* = 0$ is, to first order, the map $\frac{1}{2} g(x)$, provided $g(x)$ is defined at x.

The starlike domain of convergence $\mathcal{R}$, which lies inside the domain of invertibility $\bar{\mathcal{R}}$, is defined as follows (where $x = \rho t$ as usual):

$$\mathcal{R} := \{x = \rho t \mid t \in \mathcal{S},\ 0 < \rho < r(t)\}, \tag{31}$$

where

$$r(t) := \min \left\{ \bar{r}(t), \frac{\sigma^2(t)\hat{\rho}(s(t))}{2\delta r_b + c\sigma(t) + \sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t) \sin \hat{\phi}(s(t))}{2\delta} \right\}, \tag{32}$$

where we define

$$s(t) := \frac{g(t)}{\|g(t)\|} \in N \cap \mathcal{S},$$

and δ is the constant defined below in (39). (The coefficient 2, or $k + 1$ for the general case, in front of δr_b in the denominator of the second term of $r(t)$ is missing in [8] but is necessary for the proof of convergence.)

We conclude this subsection by characterizing the excluded directions of $\mathcal{R}$, that is, $t \in \mathcal{S}$ for which $r(t) = 0$. By the definition of $r(t)$ (32), these are directions for which at least one of $\bar{r}(t)$, $\sigma(t)$, $\|g(t)\|$, $\hat{\rho}(s(t))$, or $\sin\hat{\phi}(s(t))$ is zero. By definition, $\bar{r}(t)$ (18) is zero if and only if $\sigma(t)$ is zero. If $\sigma(t)$ is nonzero, that is, $t \notin \Pi_0^{-1}(0)$ then $g(t)$ is well defined. If additionally $\|g(t)\| \neq 0$, then $s(t)$ is well defined. Since $s(t) \in N \cap \mathcal{S}$, by (25) $\hat{\rho}(s(t))$ or $\sin\hat{\phi}(s(t))$ is zero if and only if $s(t) \in \Pi_0^{-1}(0)$. To summarize, $r(t)$ is zero for $t \in S$ if and only if one of the following conditions is true:

$$t \in \Pi_0^{-1}(0), \quad g(t) = 0, \quad \text{or} \;\; g(t)/\|g(t)\| \in \Pi_0^{-1}(0). \tag{33}$$

By the definition of Π_0 (see (15) and (10a)), the first condition fails if F satisfies 2-regularity (5) for t, and the third condition fails if F satisfies 2-regularity (5) for $g(t)/\|g(t)\|$. For the second condition, by the definition of g (30), we have for $d \in \mathbb{R}^n \setminus \Pi_0^{-1}(0)$ that

$$g(d) = 0 \Leftrightarrow \bar{B}(d)d_N + \bar{C}(d)d_{N_\perp} = 0,$$

where d_N is the orthogonal projection of d onto N and $d_{N_\perp}$ is the orthogonal projection of d onto $N_\perp$. By the definitions (10a) and (10b), we have

$$g(d) = 0 \Leftrightarrow (P_{N_*}F')'(x^*; d)d = 0, \quad \text{for } d \in \mathbb{R}^n \setminus \Pi_0^{-1}(0). \tag{34}$$

The right-hand side of this condition is identical to the condition defining the set T_2 (6), though the domain of d differs. Due to the limited smoothness of F, it is possible for either Π_0, g, or $\Pi_0(g(\cdot))$ to map a set of positive measure in $\mathbb{R}^n$ to 0. Hence, the set of excluded directions can have positive measure.

4.3 The form of a Newton step from $x \in \bar{\mathcal{R}}$

The content of this subsection is taken directly from Griewank [8] (with k set to 1); we include it here for readability of this section and for further reference in Sect. 5.

We consider the form of the Newton step from a point $x = \rho t$ in the domain of invertibility $\bar{\mathcal{R}}$ defined in (17) to the point $\bar{x}$, where

$$\bar{x} := x - F'(x)^{-1}F(x). \tag{35}$$

For $x \in \bar{\mathcal{R}}$, we have $\sigma(t) > 0$. In the remainder of this discussion, we drop the argument t from $\sigma(t)$ and the argument x from various matrix quantities such as $G(x)$, $C(x)$, etc. Using positivity of σ and (11), it can be checked that the following expressions from [8] are also true here. We have

$$F'(x)^{-1} = \begin{bmatrix} G^{-1} & -G^{-1}CE^{-1} \\ -E^{-1}DG^{-1} & E^{-1} + E^{-1}DG^{-1}CE^{-1} \end{bmatrix},$$

(see [8, (12)]), where

$$G^{-1}(x) = \rho^{-1}\bar{B}^{-1}(t) + \sigma^{-2}O(\rho^0) = \sigma^{-2}O(\rho^{-1}), \tag{36}$$

(see [8, (13)]). As in the proof of [8, Lemma 4.1] with $k = 1$, we have

$$F(x) = \begin{bmatrix} \frac{1}{2}G + O(\rho^2) & \frac{1}{2}C + O(\rho^2) \\ \frac{1}{2}D + O(\rho^2) & E + O(\rho) \end{bmatrix} x.$$

Using (11) to aggregate the order terms, as in [8], we have

$$F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I + \|G^{-1}\|O(\rho^2) & -\frac{1}{2}G^{-1}C + \|G^{-1}\|O(\rho^2) \\ O(\rho) + \|G^{-1}\|O(\rho^3) & I + O(\rho) + \|G^{-1}\|O(\rho^2) \end{bmatrix} x.$$

Due to (11), (36), and the positivity of σ, we have

$$G^{-1}(x)C(x) = \bar{B}^{-1}(t)\bar{C}(t) + \sigma^{-2}O(\rho).$$

Hence,

$$F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I + \|G^{-1}\|O(\rho^2) & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) + \sigma^{-2}O(\rho) + \|G^{-1}\|O(\rho^2) \\ O(\rho) + \|G^{-1}\|O(\rho^3) & I + O(\rho) + \|G^{-1}\|O(\rho^2) \end{bmatrix} x. \quad (37)$$

Since $\|G^{-1}\| = \sigma^{-2}O(\rho^{-1})$, we can write

$$F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & I \end{bmatrix} x - e(x), \quad (38)$$

where the remainder vector $e(x)$ can be bounded as follows:

$$\|e(x)\| \leq \delta \frac{\rho^2}{\sigma^2}, \quad (39)$$

where the constant δ is positive and finite; in fact, it is a product of finite powers of the constants in the $O(\cdot)$ terms in (11) which, as we have already noted, are finite. The definition of $r(t)$ (32) uses this value of δ.

Using (38), we have

$$\bar{x} = x - F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & \frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x + e(x) = \frac{1}{2}g(x) + e(x), \quad (40)$$

where $g(x)$ is defined in (30). In other words, if $x_k = \rho_k t_k$ for $t_k \in \mathcal{S}$ is sufficiently close to x^* and $\sigma(t_k)$ is bounded below by a positive number, then the Newton iterate x_{k+1} satisfies

$$x_{k+1} = \frac{1}{2}g(x_k) + O(\|x_k\|^2).$$

The proof provides a single positive lower bound for $\sigma(t_k)$ for all subsequent Newton iterates $\{x_k\}$. Hence, $\frac{1}{2}g(x_k)$ is a first order approximation to the Newton step from x_k.

4.4 Outline of the proof of Theorem 1

Consider the Newton iterates $\{x_j = \rho_j t_j\}_{j\geq 0}$ with $t_j \in \mathcal{S}$. For $s \in \mathcal{S}$, let $\psi_j(s)$ denote the angle between t_j and s, that is,

$$\psi_j(s) = \cos^{-1} t_j^T s. \tag{41}$$

Let $s_j = g(x_j)/\|g(x_j)\|$. The first phase of the proof is to show from (40) and the definition of $\mathcal{R}$ that if $x_0 = \rho_0 t_0 \in \mathcal{R}$, then

$$\psi_1(s_0) < \hat{\phi}(s_0) \text{ and } \rho_1 < \hat{\rho}(s_0),$$

so that $x_1 \in \mathcal{W}_{s_0}$.

The second phase of the proof analyzes convergence from inside the domain $\mathcal{W}_{s_0}$. Letting θ_j denote the angle between x_j and the null space N, it is shown that the sequence of Newton iterates $\{x_j = \rho_j t_j\}_{j\geq 1}$ starting from any point $x_1 \in \mathcal{W}_{s_0}$ maintains the properties

$$\rho_j < \hat{\rho}(s_0), \quad \theta_j < \hat{\phi}(s_0), \quad \psi_j(s_0) < \phi(s_0). \tag{42}$$

By the first and third properties, the iterates remain in $\mathcal{I}_{s_0}$ (27). Further, because of (20), the third property implies that

$$\sigma(t_j) \geq \hat{\sigma}(s_0) > 0, \tag{43}$$

a fact that is used often in the proof. Finally, it can be shown that ρ_j and θ_j go to zero as j goes to infinity and

$$\lim_{j\to\infty} \frac{\rho_{j+1}}{\rho_j} = \frac{1}{2}.$$

These facts are formally stated in Lemma 5.1 of [8]:

Lemma 1 *Suppose Assumption 1 and the standardizations (8) are satisfied. Then for any $s \in N \cap \mathcal{S} \setminus {\Pi_0}^{-1}(0)$ the Newton iteration converges linearly with common ratio $1/2$ from all points in the nonempty starlike domain $\mathcal{W}_s$. Further, the iterates remain in the starlike domain $\mathcal{I}_s$.*

Theorem 1 is obtained by combining this result with the analysis of the first step from x_0 to x_1, discussed above.

5 Acceleration of Newton's method

Overrelaxation is known to improve the rate of convergence of Newton's method to a singular solution [9]. The overrelaxed iterate is

$$x_{j+1} = x_j - \alpha F'(x_j)^{-1} F(x_j), \tag{44}$$

where α is some parameter in the range $[1, 2)$. (Of course, $\alpha = 1$ corresponds to the usual Newton step.)

We focus on a technique in which overrelaxation occurs only on every second step; that is, standard Newton steps are interspersed with steps of the form (44) for some fixed $\alpha \in [1, 2)$. Broadly speaking, each pure Newton step refocuses the error along the null space N. Kelley and Suresh [16] prove superlinear convergence for this method when α is systematically increased to 2 as the iterates converge. However, their proof requires the third derivative of F evaluated at x^* to satisfy a boundedness condition.

In this section, we state our main result and motivate its proof, highlighting some key points. The lengthy proof appears in full in [18, Sect. 5].

We assume that 2^1-regularity holds at x^*, F' is strongly semismooth at x^*, and that $x_0 \in \mathcal{R}_\alpha$, where $\mathcal{R}_\alpha$ is a starlike domain defined in (60) whose excluded directions are identical to those of $\mathcal{R}$ defined in Sect. 4 but whose rays are shorter. In fact, as α is increased to 2, the rays of the starlike domain $\mathcal{R}_\alpha$ shrink in length to zero.

Theorem 2 *Suppose Assumption 1 holds and let $\alpha \in [1, 2)$. There exists a starlike domain $\mathcal{R}_\alpha \subseteq \mathcal{R}$ about x^* such that if $x_0 \in \mathcal{R}_\alpha$ and with iterates defined by*

$$x_{2j+1} = x_{2j} - F'(x_{2j})^{-1} F(x_{2j}) \quad \textit{and} \tag{45}$$

$$x_{2j+2} = x_{2j+1} - \alpha F'(x_{2j+1})^{-1} F(x_{2j+1}), \tag{46}$$

for $j = 0, 1, 2, \ldots$, then the iterates $\{x_i\}$ for $i = 0, 1, 2, \ldots$ converge linearly to x^* and

$$\lim_{j\to\infty} \frac{\|x_{2j+2} - x^*\|}{\|x_{2j} - x^*\|} = \frac{1}{2}\left(1 - \frac{\alpha}{2}\right).$$

We first describe a key step of the proof. Since the problem is in standard form, we have from (40) that the Newton step (45) satisfies the following relationships for $x_{2k} \in \bar{\mathcal{R}}$:

$$x_{2k+1} = \frac{1}{2}\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + e(x_{2k}) = \frac{1}{2} g(x_{2k}) + e(x_{2k}), \tag{47}$$

for all $k \geq 0$, where $g(\cdot)$ is defined in (30) and the remainder term $e(\cdot)$ is defined in (38). As in (39), we have

$$\|e(x_{2k})\| \leq \delta \frac{\rho_{2k}^2}{\sigma_{2k}^2}. \tag{48}$$

For the accelerated Newton step (46), using manipulations similar to those leading to (40), we have for $x_{2k+1} \in \bar{\mathcal{R}}$ that

$$x_{2k+2} = \begin{bmatrix} (1 - \frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix} x_{2k+1} + \alpha e(x_{2k+1}), \tag{49}$$

for all $k \geq 0$. By substituting (47) into (49), we obtain

$$x_{2k+2} = \tfrac{1}{2}\begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix}\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + \tilde{e}_\alpha(x_{2k}, x_{2k+1}), \tag{50}$$

where

$$\tilde{e}_\alpha(x_{2k}, x_{2k+1}) = \begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix} e(x_{2k}) + \alpha e(x_{2k+1}). \tag{51}$$

Multiplying the matrices in (50), we have

$$\begin{aligned} x_{2k+2} &= \frac{1}{2}\left(1-\frac{\alpha}{2}\right)\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + \tilde{e}_\alpha(x_{2k}, x_{2k+1}) \\ &= \frac{1}{2}\left(1-\frac{\alpha}{2}\right) g(x_{2k}) + \tilde{e}_\alpha(x_{2k}, x_{2k+1}), \end{aligned} \tag{52}$$

To bound the remainder term, note that $|1-\frac{\alpha}{2}| + |1-\alpha| = \frac{\alpha}{2}$ for $\alpha \in [1, 2)$, so we have from (51) that

$$\begin{aligned} \|\tilde{e}_\alpha(x_{2k}, x_{2k+1})\| &\leq \frac{\alpha}{2}\left(1 + \|\bar{B}(t_{2k+1})^{-1}\|\|\bar{C}(t_{2k+1})\|\right)\|e(x_{2k})\| + \alpha\|e(x_{2k+1})\| \\ &\leq \left(\frac{\sigma_{2k+1} + \|\bar{C}(t_{2k+1})\|}{\sigma_{2k+1}}\right)\delta\frac{\rho_{2k}^2}{\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\qquad \text{from } \alpha < 2, \text{(14), and (39)} \\ &\leq c\delta\frac{\rho_{2k}^2}{\sigma_{2k+1}\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\qquad \text{from (21)} \\ &\leq \tilde{\delta}\frac{\rho_{2k}^2 + \rho_{2k+1}^2}{\mu_{2k}^3}, \end{aligned} \tag{53}$$

where

$$\mu_{2k} := \min(\sigma_{2k}, \sigma_{2k+1}) \tag{54}$$

and $\tilde{\delta} := \delta\max(c, \alpha)$. If $x_{2k} = \rho_{2k}t_{2k}$ for $t_{2k} \in \mathcal{S}$ and $x_{2k+1} = \rho_{2k+1}t_{2k+1}$ for $t_{2k+1} \in \mathcal{S}$ are sufficiently close to x^* and $\sigma(t_{2k})$ and $\sigma(t_{2k+1})$ are bounded below by a positive number, then x_{2k+2} satisfies

$$x_{2k+2} = \frac{1}{2}(1-\frac{\alpha}{2})g(x_{2k}) + O(\|x_{2k}\|^2).$$

The proof provides a single positive lower bound for $\sigma(t_{2k})$ and $\sigma(t_{2k+1})$ for all subsequent iterates. Hence, $\frac{1}{2}\left(1-\frac{\alpha}{2}\right)g(x_{2k})$ is a first order approximation to the double step achieved by applying a Newton step followed by an overrelaxed Newton step from x_{2k}.

We introduce the following new parameters:

$$q_\alpha(s) := \frac{1-\alpha/2}{4}\sin\phi(s), \text{ for } s\in N\cap\mathcal{S}, \tag{55}$$

We define the angle $\tilde{\phi}_\alpha(s)$, for which $0\le\tilde{\phi}_\alpha(s)\le\pi/2$, by the equality

$$\sin\tilde{\phi}_\alpha(s) := \min\left\{\frac{q_\alpha(s)}{c/\hat{\sigma}(s)+1-q_\alpha(s)}, \frac{2\delta\hat{r}(s)}{(1-q_\alpha(s))\hat{\sigma}^2(s)}\right\}, \text{ for } s\in N\cap\mathcal{S}. \tag{56}$$

We further define

$$\tilde{\rho}_\alpha(s) := \frac{(1-\alpha/2-q_\alpha(s))\hat{\sigma}^3(s)}{4\tilde{\delta}}\sin\tilde{\phi}_\alpha(s) \text{ for } s\in N\cap\mathcal{S}, \tag{57}$$

$$\mathcal{W}_{s,\alpha} := \{x=\rho t\mid t\in\mathcal{S},\ \cos^{-1}(t^Ts)<\tilde{\phi}_\alpha(s),\ 0<\rho<\tilde{\rho}_\alpha(s)\}, \tag{58}$$

and

$$\mathcal{I}_{s,\alpha} := \{x=\rho t\mid t\in\mathcal{S},\ \cos^{-1}(t^Ts)<\phi(s),\ 0<\rho<\tilde{\rho}_\alpha(s)\}. \tag{59}$$

It can be shown that $\mathcal{W}_{s,\alpha}\subseteq\mathcal{I}_{s,\alpha}\subseteq\bar{\mathcal{R}}$. The starlike domain of convergence is defined as follows:

$$\mathcal{R}_\alpha := \{x=\rho t\mid t\in\mathcal{S},\ 0<\rho<r_\alpha(t)\}, \tag{60}$$

where

$$r_\alpha(t) := \min\left\{\bar{r}(t), \frac{\sigma^2(t)\tilde{\rho}_\alpha(s(t))}{2\delta r_b+c\sigma(t)+\sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t)(1-\alpha/2)\sin\tilde{\phi}_\alpha(s(t))}{8\delta}\right\} \tag{61}$$

and $s(t)=g(t)/\|g(t)\|\in N\cap\mathcal{S}$.

As in Sect. 4, the angle between iterate $x_i=\rho_i t_i$ and the null space N is denoted by θ_i, while $\psi_i(s_0)$ denotes the angle between x_i and s_0 (41). The proof of Theorem 2 is by induction. The induction step consists of showing that if

$$\rho_{2k+\iota}<\tilde{\rho}_\alpha(s_0),\quad \theta_{2k+\iota}<\tilde{\phi}_\alpha(s_0),\ \text{and } \psi_{2k+\iota}(s_0)<\phi(s_0),$$
$$\text{for } \iota\in\{1,2\},\quad \text{all } k \text{ with } 0\le k<j, \tag{62}$$

then

$$\rho_{2j+\iota}<\tilde{\rho}_\alpha,\quad \theta_{2j+\iota}<\tilde{\phi}_\alpha(s_0),\ \text{and } \psi_{2j+\iota}(s_0)<\phi(s_0)\quad \text{for } \iota\in\{1,2\}. \tag{63}$$

For all $i=1,2,\ldots,$ the third property in (62) and (63), $\psi_i(s_0)<\phi(s_0)$, implies that $\sigma(t_i)\ge\hat{\sigma}(s_0)>0$; see (20) and (43). By the first and third properties, the iterates remain in $\mathcal{I}_{s_0,\alpha}$. Since $\mathcal{I}_{s_0,\alpha} \subseteq \bar{\mathcal{R}}$, the bounds of Subsect. 4.3 together with (47) and (49) are valid for our iterates.

The anchor step of the induction argument consists of showing that for $x_0 \in \mathcal{R}_\alpha$, we have $x_1 \in \mathcal{W}_{s_0,\alpha}$ and $x_2 \in \mathcal{I}_{s_0,\alpha}$ with $\theta_2 < \tilde{\phi}_\alpha$. Indeed, these facts yield (62) for $j = 1$.

The convergence rate claimed in the theorem is a byproduct of the proof of the induction step.

6 Application to nonlinear complementarity problems

The nonlinear complementarity problem for the function $f : \mathbb{R}^n \to \mathbb{R}^n$ is as follows: Find an $x \in \mathbb{R}^n$ such that

$$0 \leq f(x), \quad x \geq 0, \quad x^T f(x) = 0. \qquad \text{NCP}(f)$$

Let x^* be a solution of NCP(f). We assume that f' is well defined and strongly semismooth at x^*. We apply a nonlinear-equations reformulation to the NCP. We do not standardize the resulting equations (as we did earlier in (8) to simplify the discussions of Sects. 4 and 5), as the rescaling and shifting needed to enforce this assumption would complicate this section considerably.

We tailor the convergence results of previous sections to this reformulation, interpret the 2-regularity condition for the NCP(f), and provide conditions under which the starlike domain of convergence is "directionally dense" at the solution.

6.1 NCP notation

We use e_i to denote the ith column of the identity matrix. The notation $\langle\cdot,\cdot\rangle$ denotes the inner product between two vectors. For any $x \in \mathbb{R}^n$, we use diag x to denote the $\mathbb{R}^{n\times n}$ diagonal matrix formed from the components of x.

We define the inactive, biactive, and active index sets, α, β, and γ respectively, at a solution x^* of NCP(f) as follows,

$$\begin{cases} i \in \alpha, \text{ if } x_i^* = 0, \ \ f_i(x^*) > 0, \\ i \in \beta, \text{ if } x_i^* = 0, \ \ f_i(x^*) = 0, \\ i \in \gamma, \text{ if } x_i^* > 0, \ \ f_i(x^*) = 0. \end{cases}$$

For the vector function f and its derivatives, we give a sample of the notational conventions used below. $f_\gamma(x^*)$ denotes the $|\gamma|$-vector whose components are $f_i(x^*)$, $i \in \gamma$. We use $f'_{i,\gamma}(x^*)$to denote the vector in $\mathbb{R}^{|\gamma|}$ with elements $\frac{df_i}{dx_j}(x^*)$, $j \in \gamma$, while $f'_{\gamma,\alpha}(x^*)$ denotes the matrix in $\mathbb{R}^{|\gamma|\times|\alpha|}$ whose elements are $\frac{df_i}{dx_j}(x^*)$, for $i \in \gamma$ and $j \in \alpha$. The notation $f'_\gamma(x^*)$ represents the matrix in $\mathbb{R}^{|\gamma|\times n}$ whose elements are $\frac{df_i}{dx_j}(x^*)$, for $i \in \gamma$ and $j = 1, 2, \ldots, n$.

6.2 The nonlinear-equations reformulation

Recall the nonlinear-equations reformulation Ψ (7) of the NCP (1), and consider the use of Newton's method for solving $\Psi(x) = 0$. Taking the derivative of Ψ, we have

$$\begin{aligned}\Psi_i'(x) = 2\{&(f_i(x) - \min(0, x_i + f_i(x)))e_i \\ &+ (x_i - \min(0, x_i + f_i(x)))f_i'(x)\}, \quad \text{for } i = 1, 2, \ldots, n. \end{aligned} \tag{64}$$

It can be seen that Ψ' is strongly semismooth when f' is strongly semismooth by applying the following two facts: From [6, Proposition 7.4.4], the composition of strongly semismooth functions is strongly semismooth, and from [6, Proposition 7.4.7], every piecewise-affine map is strongly semismooth.

At the solution x^*, Ψ_i' simplifies to

$$\Psi_i'(x^*) = 2\{f_i(x^*)e_i + x_i^* f_i'(x^*)\}.$$

By inspection, we have

$$\begin{cases} \Psi_i'(x^*) = 2f_i(x^*)e_i, & i \in \alpha, \\ \Psi_i'(x^*) = 0, & i \in \beta, \\ \Psi_i'(x^*) = 2x_i^* f_i'(x^*), & i \in \gamma. \end{cases}$$

The null space of $\Psi'(x^*)$ (whose ith row is the transpose of Ψ_i') is

$$N \equiv \ker \Psi'(x^*) = \{\xi \in \mathbb{R}^n \mid f_\gamma'(x^*)\xi = 0,\ \xi_\alpha = 0\}, \tag{65}$$

so that

$$\dim N = \dim \ker f'_{\gamma,\beta\cup\gamma}(x^*).$$

In particular, if $\beta \neq \emptyset$, then $\dim N > 0$ and x^* is a singular solution of $\Psi(x) = 0$. The null space of $\Psi'(x^*)^T$ is

$$\begin{aligned} N_* = \{\xi \in \mathbb{R}^n \mid \xi_\alpha = -(\operatorname{diag} f_\alpha(x^*))^{-1}(f'_{\gamma,\alpha}(x^*))^T(\operatorname{diag} x_\gamma^*)\xi_\gamma, \\ f'_{\gamma,\beta\cup\gamma}(x^*)^T(\operatorname{diag} x_\gamma^*)\xi_\gamma = 0\}. \end{aligned} \tag{66}$$

If $\operatorname{rank} f'_{\gamma,\beta\cup\gamma}(x^*) = |\gamma|$, then $N_* = \{\xi \in \mathbb{R}^n \mid \xi_\alpha = 0,\ \xi_\gamma = 0\}$.

The 2-regularity condition (5) for Ψ at x^* and $d \in \mathbb{R}^n$ is

$$(P_{N_*}\Psi')'(x^*; d)|_N \ \text{ is nonsingular.} \tag{67}$$

By direct calculation, we have

$$\begin{aligned}\frac{1}{2}(\Psi')_i'(x; d) = (\langle f_i'(x), d\rangle - \eta_i)e_i + (d_i - \eta_i)f_i'(x) \\ + (x_i - \min(0, x_i + f_i(x)))(f_i')'(x; d),\end{aligned}$$

where $\eta_i := \min(0, x_i + f_i(x))'(x; d)$. We calculate this quantity using the result [6, Proposition 3.1.6] for the composition of B-differentiable functions:

$$\eta_i = \begin{cases} \min(0, d_i + \langle f_i'(x), d\rangle), & \text{if } x_i + f_i(x) = 0, \\ 0, & \text{if } x_i + f_i(x) > 0, \\ d_i + \langle f_i'(x), d\rangle, & \text{if } x_i + f_i(x) < 0. \end{cases}$$

At a solution x^*, we have $\eta_i = 0$ for $i \in \alpha \cup \gamma$, and $\eta_i = \min(0, d_i + \langle f_i'(x^*), d\rangle)$ for $i \in \beta$. Hence, we have

$$\frac{1}{2}(\Psi_i')'(x^*; d) = \begin{cases} \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*), & i \in \alpha, \\ (\langle f_i'(x^*), d\rangle - \min(0, d_i + \langle f_i'(x^*), d\rangle))e_i \\ \quad + (d_i - \min(0, d_i + \langle f_i'(x^*), d\rangle)) f_i'(x^*), & i \in \beta, \\ \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*) + x_i^*(f_i')'(x^*; d), & i \in \gamma. \end{cases} \tag{68}$$

By noting that for any scalars s_1, s_2 we have

$$s_1 - \min(0, s_2) = s_1 + \max(0, -s_2) = \max(s_1, s_1 - s_2) = -\min(-s_1, s_2 - s_1),$$

we can rewrite (68) as follows

$$\frac{1}{2}(\Psi_i')'(x^*; d) = \begin{cases} \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*), & i \in \alpha, \\ \max(\langle f_i'(x^*), d\rangle, -d_i)e_i - \min(\langle f_i'(x^*), d\rangle, -d_i) f_i'(x^*), & i \in \beta, \\ \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*) + x_i^*(f_i')'(x^*; d), & i \in \gamma. \end{cases} \tag{69}$$

Using the notation

$$r = \operatorname{rank} f'_{\gamma,\beta\cup\gamma}(x^*),$$

we define an orthonormal matrix Z of dimension $|\gamma| \times r$ such that the columns of Z span range $f'_{\gamma,\beta\cup\gamma}(x^*)$, and another orthonormal matrix $Z_\perp$ of dimensions $|\gamma| \times (|\gamma| - r)$ such that the columns of $Z_\perp$ span $\ker f'_{\gamma,\beta\cup\gamma}(x^*)^T$. Note that $[Z \mid Z_\perp]$ is an orthogonal matrix of dimensions $|\gamma| \times |\gamma|$. (The matrices Z and $Z_\perp$ are not uniquely defined by the conditions above, but the properties discussed below are independent of the particular choices used).

In the remainder of this section, we often drop the argument x^* from f and f', for clarity.

Proposition 1 *2-regularity (67) holds for $d \in \mathbb{R}^n$ at a solution x^* of $\Psi(x) = 0$ if and only if the matrix*

$$\begin{bmatrix} [e_i^T]_{i\in\alpha} \\ [\max(\langle f_i', d\rangle, -d_i)e_i - \min(\langle f_i', d\rangle, -d_i) f_i']_{i\in\beta}^T \\ Z^T f_\gamma' \\ Z_\perp^T \left[(f_i')'(x^*; d) + (1/x_i^*)\langle f_i', d\rangle e_i^T - f_{i,\alpha}' \operatorname{diag}(d_j/f_j)_{j\in\alpha} f_\alpha'\right]_{i\in\gamma} \end{bmatrix} \tag{70}$$

is nonsingular. Further, for $d \in N$, 2-regularity holds if and only if the simpler matrix

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\max(\langle f_i', d\rangle, -d_i)e_i - \min(\langle f_i', d\rangle, -d_i) f_i'\right]_{i\in\beta}^T \\ Z^T f_\gamma' \\ Z_\perp^T (f_\gamma')'(x^*; d) \end{bmatrix} \tag{71}$$

is nonsingular.

Proof The claim that $(P_{N_*}\Psi')'(x^*; d)|_N$ is nonsingular for some $d \in \mathbb{R}^n$ (67) is equivalent to

$$P_{N_*}(\Psi')'(x^*; d)v = 0 \text{ and } v \in N \;\Rightarrow\; v = 0.$$

For $v \in N$, we have from (65) and (69) that

$$\frac{1}{2}(\Psi_i')'(x^*; d)v = \begin{cases} d_i\langle f_i', v\rangle, & i \in \alpha \\ \max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle, & i \in \beta \\ \langle f_i', d\rangle v_i + x_i^*\langle (f_i')'(x^*; d), v\rangle, & i \in \gamma. \end{cases} \tag{72}$$

Since N_* is defined in (66) to have the form $\{\xi \in \mathbb{R}^n \mid A\xi = 0\}$ for some matrix A, we have that $P_{N_*}w = 0$ if and only if $w = A^T z$ for some z. In our case, we have

$$\frac{1}{2}(\Psi')'(x^*; d)v = \begin{bmatrix} \operatorname{diag} f_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ (\operatorname{diag} x_\gamma^*) f_{\gamma,\alpha}' & (\operatorname{diag} x_\gamma^*) f_{\gamma,\beta}' & (\operatorname{diag} x_\gamma^*) f_{\gamma,\gamma}' \end{bmatrix} \begin{bmatrix} z_\alpha \\ z_\beta \\ z_\gamma \end{bmatrix}, \tag{73}$$

for some $z \in \mathbb{R}^n$. By matching components from this expression and (72), we have that $P_{N_*}(\Psi')'(x^*; d)v = 0$ if for some $z \in \mathbb{R}^n$ we have

$$\begin{aligned} &d_i\langle f_i', v\rangle = z_i f_i, && i \in \alpha, \\ &\max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle = 0, && i \in \beta, \\ &\langle f_i', d\rangle v_i + x_i^*\langle (f_i')'(x^*; d), v\rangle = x_i^*\left[\langle f_{i,\alpha}', z_\alpha\rangle + \langle f_{i,\beta}', z_\beta\rangle + \langle f_{i,\gamma}', z_\gamma\rangle\right], && i \in \gamma. \end{aligned}$$

Rearranging the first equation above yields an expression for z_α, which can be substituted into the third equation to give the following:

$$0 = \max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle, \quad i \in \beta, \tag{74a}$$

$$\begin{aligned} &\langle f_i', d\rangle v_i + x_i^*\langle (f_i')'(x^*; d), v\rangle - x_i^*\langle f_{i,\alpha}', \operatorname{diag}\left(d_j/f_j\right)_{j\in\alpha} f_\alpha' v\rangle \\ &\quad = x_i^*\left[f_{i,\beta}' z_\beta + f_{i,\gamma}' z_\gamma\right], \quad i \in \gamma. \end{aligned} \tag{74b}$$

Using the definition of Z, we can rewrite (74b) as follows:

$$\left[(1/x_i^*)\langle f_i', d\rangle v_i + \langle (f_i')'(x^*; d), v\rangle - \langle f_{i,\alpha}', \operatorname{diag}\left(d_j/f_j\right)_{j\in\alpha} f_\alpha' v\rangle\right]_{i\in\gamma} = Zt,$$

for some $t \in \mathbb{R}^r$, so that

$$Z_\perp^T \left[(1/x_i^*)\langle f_i', d\rangle e_i^T + (f_i')'(x^*; d) - f_{i,\alpha}', \operatorname{diag}\left(d_j/f_j\right)_{j\in\alpha} f_\alpha'\right]_{i\in\gamma} v = 0. \tag{75}$$

Since $v \in N$, we have from (65) that

$$v_\alpha = 0, \qquad f_\gamma' v = 0. \tag{76}$$

The second condition of (76) is equivalent to

$$\begin{bmatrix} Z^T \\ Z_\perp^T \end{bmatrix} f_\gamma' v = \begin{bmatrix} Z^T \\ Z_\perp^T \end{bmatrix} \left[f_{\gamma,\alpha}' \; f_{\gamma,\beta}' \; f_{\gamma,\gamma}'\right] v = 0. \tag{77}$$

Because

$$Z_\perp^T \left[f_{\gamma,\alpha}' \; f_{\gamma,\beta}' \; f_{\gamma,\gamma}'\right] v = \left[Z_\perp^T f_{\gamma,\alpha}' \; 0 \; 0\right] v = Z_\perp^T f_{\gamma,\alpha}' v_\alpha$$

and $v_\alpha = 0$, the second block row in the system (77) does not add any information and can be dropped. Hence, we can write (76) equivalently as

$$v_\alpha = 0, \qquad Z^T f_\gamma' v = 0. \tag{78}$$

By gathering the conditions equivalent to $v \in N$ and $P_{N_*}(\Psi')'(x^*; d)v = 0$, namely, (74a), (75), and (78), we have

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\max(\langle f_i', d\rangle, -d_i)e_i - \min(\langle f_i', d\rangle, -d_i) f_i'\right]_{i\in\beta}^T \\ Z^T f_\gamma' \\ Z_\perp^T \left[(f_i')'(x^*; d) + (1/x_i^*)\langle f_i', d\rangle e_i^T - f_{i,\alpha}' \operatorname{diag}\left(d_j/f_j\right)_{j\in\alpha} f_\alpha'\right]_{i\in\gamma} \end{bmatrix} v = 0,$$

from which we deduce that $v = 0$ whenever the coefficient matrix in this expression is nonsingular. Hence x^* is 2-regular for Ψ with respect to $d \in \mathbb{R}^n$ if the matrix (70) is nonsingular. For $d \in N$, we have by the definition of N (65) that $\langle f_i', d\rangle = 0$ for $i \in \gamma$ and $d_\alpha = 0$. Upon applying these simplifications to the above matrix, we have precisely the matrix (71). □

Recall from Definition 6 that Ψ (7) is 2^1-regular at x^* if $(P_{N_*}\Psi')'(x^*; d)|_N$ is nonsingular for some d in N, that is, if the matrix (71) is nonsingular for some $d \in N$. The following theorem specializes Theorems 1 and 2 for applying Newton's method to the nonlinear-equations reformulation $\Psi(x)$ of NCP(f).

Theorem 3 *Consider a solution x^* of NCP(f) for $f : \mathbb{R}^n \to \mathbb{R}^n$ with f' strongly semismooth at x^*. Suppose that x^* is a singular solution in the sense that $N = \ker f'_{\gamma,\beta\cup\gamma}(x^*)$ is nontrivial. Suppose also that the matrix (71) is nonsingular for some $d \in N$. Then there exists a starlike domain $\mathcal{R}$ about x^*, such that, if Newton's method for the nonlinear-equations reformulation $\Psi(x)$ is initialized at any $x_0 \in \mathcal{R}$, the iterates converge linearly to x^* with rate $1/2$. Furthermore, if Newton's method is accelerated according to (45) and (46) for some $\alpha \in [1, 2)$, then there exists a starlike domain $\mathcal{R}_\alpha \subseteq \mathcal{R}$ about x^*, such that if $x_0 \in \mathcal{R}_\alpha$ then the accelerated iterates $\{x_i\}$ for $i = 0, 1, 2, \ldots,$ converge linearly to x^* and*

$$\lim_{j\to\infty} \frac{\|x_{2j+2} - x^*\|}{\|x_{2j} - x^*\|} = \frac{1}{2}\left(1 - \frac{\alpha}{2}\right).$$

6.3 2-regularity conditions for special cases of the NCP

In this section we show that the regularity conditions (70) and (71) simplify to more familiar regularity conditions in special cases of the NCP.

Nondegenerate NCP. Consider the case of nondegenerate NCP. We obtain a simpler regularity condition, related to 2-regularity for nonlinear equations, that ensures that 2-regularity holds for some $d \in N$, and hence that the conditions of Theorem 3 are satisfied.

Theorem 4 *Suppose that $\beta = \emptyset$. Then the NCP satisfies 2^1-regularity at the solution x^* if and only if*

$$P_{N^f_{*\gamma}} (f'_{\gamma,\gamma})'(x^*; d)|_{N^f_\gamma} \tag{79}$$

is nonsingular for $d \in N$, where

$$N^f_\gamma = \{\xi_\gamma \in \mathbb{R}^{|\gamma|} \mid f'_{\gamma,\gamma}\xi_\gamma = 0\}, \qquad N^f_{*\gamma} = \{\xi_\gamma \in \mathbb{R}^{|\gamma|} \mid (f'_{\gamma,\gamma})^T\xi_\gamma = 0\}.$$

Proof Let the orthonormal matrices $Z_\perp$ and Z be as in (71), and define two additional orthonormal matrices $\bar{Z}$ and $\bar{Z}_\perp$ such that the columns of $\bar{Z}_\perp$ span $\ker f'_{\gamma,\gamma}$ (and hence the space N^f_γ), the columns of $\bar{Z}$ span range $(f'_{\gamma,\gamma})^T$, and $\left[\bar{Z} \mid \bar{Z}_\perp\right]$ is orthogonal. We have $\bar{Z} \in \mathbb{R}^{|\gamma|\times r}$ and $\bar{Z}_\perp \in \mathbb{R}^{|\gamma|\times(|\gamma|-r)}$. Specializing 2-regularity for $d \in N$ (71) to the case of $\beta = \emptyset$, we have that 2^1-regularity is equivalent to nonsingularity of the following matrix for some $d \in N$:

$$\begin{bmatrix} [e_i^T]_{i\in\alpha} \\ Z^T\left[f'_{\gamma,\alpha}(x^*)\ f'_{\gamma,\gamma}(x^*)\right] \\ Z_\perp^T(f'_\gamma)'(x^*; d) \end{bmatrix} \begin{bmatrix} I_\alpha & 0 \\ 0 & [\bar{Z}\ \bar{Z}_\perp] \end{bmatrix},$$

where I_α is the identity matrix of dimension $|\alpha|$. By forming the matrix product, we find that it is block lower triangular. Therefore, nonsingularity of the matrix product is equivalent to nonsingularity of the three (square) diagonal blocks, which are

$$I_\alpha, \qquad Z^T f'_{\gamma,\gamma}(x^*)\bar{Z}, \qquad Z_\perp^T (f'_{\gamma,\gamma})'(x^*; d)\bar{Z}_\perp,$$

which have dimensions $|\alpha|$, r, and $|\gamma|-r$, respectively. It is easy to see that $Z^T f'_{\gamma,\gamma}(x^*)$ $\bar{Z}$ is nonsingular by the definition of Z and $\bar{Z}$. Since the columns of $Z_\perp$, as defined earlier, must span the subspace $N^f_{*\gamma}$, and since the columns of $\bar{Z}_\perp$ span the subspace N^f_γ, nonsingularity of $Z_\perp^T (f'_{\gamma,\gamma})'(x^*; d)\bar{Z}_\perp$ is equivalent to condition (79). □

Nonlinear equations. We now consider the case in which $\alpha = \beta = \emptyset$, so that the NCP reduces essentially to a system of nonlinear equations $f(x) = 0$ whose solution is at $x = x^*$. In the nondegenerate case in which $f'_{\gamma,\gamma}(x^*) \equiv f'(x^*)$ has full rank n, we have from definition (65) that $N = \{0\}$, so that x^* is a nonsingular solution and Theorem 3 does not apply. Therefore, suppose that $f'(x^*)$ has rank less than n and $\alpha = \beta = \emptyset$—essentially the case of singular nonlinear equations. By specializing the discussion of nondegenerate NCP, we have from the definitions in Theorem 4 that

$$N^f = \ker f'(x^*), \qquad N^f_* = \ker f'(x^*)^T,$$

where we have dropped the subscript γ. Hence, 2-regularity is satisfied for some $d \in N$ if

$$P_{N^f_*}(f')'(x^*; d)|_{N^f} \qquad \text{is nonsingular for some } d \in N.$$

This is the 2^1-regularity condition for nonlinear equations (Definition 6).

NCP with a modified weak regularity condition. We now consider another special case in which we remove the condition $\beta = \emptyset$ and assume that the matrix $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank. This assumption is similar to the weak regularity condition of Daryina et al. [1], which is a full-rank assumption on $f'_{\beta\cup\gamma,\gamma}(x^*)$. (The two assumptions are identical when $\beta = \emptyset$ or f' is symmetric, as is the case when f is the gradient of a scalar function).

Theorem 5 *If for $d \in \mathbb{R}^n$ the set of n vectors in $\mathbb{R}^n$*

$$\begin{aligned} \{e_i\}_{i\in\alpha} \cup \{f'_i(x^*)\}_{i\in\gamma} \cup \{\langle f'_i(x^*), d\rangle e_i + d_i f'_i(x^*)\}_{i\in\beta_1} \\ \cup \{\langle f'_i(x^*), d\rangle f'_i(x^*) + d_i e_i\}_{i\in\beta_2}, \end{aligned} \tag{80}$$

where $\beta_1 := \beta_1(d)$ and $\beta_2 := \beta_2(d)$, with

$$\beta_1(d) := \{i \in \beta \mid \langle f'_i(x^*), d\rangle > -d_i\}, \tag{81a}$$

$$\beta_2(d) := \{i \in \beta \mid \langle f'_i(x^*), d\rangle \le -d_i\}, \tag{81b}$$

is linearly independent, then 2-regularity (70) is satisfied by the NCP at x^* for $d \in \mathbb{R}^n$. Conversely, if $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank and 2-regularity holds for $d \in \mathbb{R}^n$ at x^*, then the set of vectors (80) is linearly independent.

Proof Observe that if $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank, we can set $Z = I$ and $Z_\perp$ null, so the matrix in (70) reduces to

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\max(\langle f_i'(x^*), d\rangle, -d_i)e_i - \min(\langle f_i'(x^*), d\rangle, -d_i) f_i'(x^*)\right]_{i\in\beta}^T \\ f'_\gamma(x^*) \end{bmatrix}.$$

By partitioning the index set β according to (81), we see that nonsingularity of this matrix is equivalent to linear independence of the vectors (80). □

As discussed at the end of Sect. 4, 2-regularity for almost every $d \in \mathbb{R}^n$ is necessary for "directional denseness" of the starlike domain of convergence. According to Theorem 5, it is sufficient to require linear independence of the vectors (80) for the partition (β_1, β_2) of β defined in (81) for almost every $d \in \mathbb{R}^n$. This condition is similar to the quasi-regularity condition of Izmailov and Solodov [12, Definition 4.1], which requires linear independence of the vectors (80) for every partition (β_1, β_2) of β for some fixed $d \in \mathbb{R}^n$.

6.4 "Directional denseness" of the starlike domain

In this subsection, we give sufficient conditions for the starlike domain of convergence $\mathcal{R}$ (31) (or $\mathcal{R}_\alpha$ (60)), to be "directionally dense" at the solution x^* in terms of f.

Definition 8 A starlike domain $\mathcal{R}$ about $x^* \in \mathbb{R}^n$ is *directionally dense* at x^* if for almost every $t \in \mathcal{S}$,

$$\text{there exists } C_t > 0 \text{ such that } x = x^* + \rho t \in \mathcal{R} \text{ for all } \rho \in (0, C_t). \tag{82}$$

A direction t satisfies (82) if and only if t is not an *excluded direction*, as defined in Sect. 2.

We recall the characterization of the excluded directions of $\mathcal{R}$ from (33): A direction $t \in \mathcal{S}$ is excluded if and only if one of the following is true:

$$t \in \Pi_0^{-1}(0), \quad g(t) = 0, \quad \text{or} \quad g(t)/\|g(t)\| \in \Pi_0^{-1}(0). \tag{83}$$

The first condition of (83) fails if Ψ satisfies the 2-regularity condition (67) for t, and the third condition of (83) fails if Ψ satisfies the 2-regularity condition (67) for $g(t)/\|g(t)\|$. Applying Proposition 1 and noting that range $g = N$, the first condition of (83) fails if the matrix (70) is nonsingular for $d = t$ and the third condition of (83) fails if the simpler matrix (71) is nonsingular for $d = g(t)/\|g(t)\|$.

Now consider the second condition of (83). For $x \in \mathbb{R}^n$ with $\Pi_0(x - x^*) \neq 0$ and $\|x - x^*\|$ sufficiently small, recall from (30) that the Newton iterate from x is

$x^* + \frac{1}{2}g(x - x^*) + O(\|x - x^*\|^2)$, where $g : (\mathbb{R}^n \setminus \Pi_0^{-1}(0)) \to N \subseteq \mathbb{R}^n$ is the positively homogeneous vector defined by

$$g(x - x^*) = \rho g(t) = P_N(x - x^*) \\ + ((P_{N_*}\Psi')'(x^*; t)|_N)^{-1}(P_{N_*}\Psi')'(x^*; t)|_{N_\perp} P_{N_\perp}(x - x^*), \quad (84)$$

for $x = x^* + \rho t$, $\rho = \|x - x^*\|$, and $t \in \mathcal{S}$. As in (34), we have

$$g(d) = 0 \Leftrightarrow (P_{N_*}\Psi')'(x^*; d)d = 0, \qquad \text{for } d \in \mathbb{R}^n \setminus \Pi_0^{-1}(0). \quad (85)$$

From (69), dividing the set β into $\beta_1(d)$ and $\beta_2(d)$ (81) for $d \in \mathbb{R}^n$, we have

$$\frac{1}{2}(\Psi_i')'(x^*; d)d = \begin{cases} 2d_i\langle f_i'(x^*), d\rangle, & i \in \alpha, \\ 2d_i\langle f_i'(x^*), d\rangle, & i \in \beta_1(d), \\ -d_i^2 - \langle f_i'(x^*), d\rangle^2, & i \in \beta_2(d), \\ 2d_i\langle f_i'(x^*), d\rangle + x_i^*\langle (f_i')'(x^*; d), d\rangle, & i \in \gamma. \end{cases} \quad (86)$$

To express $(P_{N_*}\Psi')'(x^*; d)d = 0$ in terms of f, we recall from the proof of Proposition 1 that $P_{N_*}w = 0$ if and only if $w = A^T z$ for some $z \in \mathbb{R}^n$, where $A^T z$ is the right-hand side of (73). That is, $(P_{N_*}\Psi')'(x^*; d)d = 0$ for $d \in \mathbb{R}^n$ if and only if there is some $z \in \mathbb{R}^n$ for which

$$2d_i\langle f_i'(x^*), d\rangle = f_i z_i, \qquad i \in \alpha, \quad (87a)$$
$$2d_i\langle f_i'(x^*), d\rangle = 0, \qquad i \in \beta_1(d), \quad (87b)$$
$$d_i^2 + \langle f_i'(x^*), d\rangle^2 = 0, \qquad i \in \beta_2(d), \quad (87c)$$
$$2d_i\langle f_i'(x^*), d\rangle + x_i^*\langle (f_i')'(x^*; d), d\rangle = x_i^*\langle f_i', z\rangle, \qquad i \in \gamma. \quad (87d)$$

Thus, if $t \in \mathbb{R}^n \setminus \Pi_0^{-1}(0)$ and (87) has no solution z for $t = d$, then $g(t) \neq 0$ and the second condition of (83) fails. In fact, it seems quite likely that (87) has no solution $z \in \mathbb{R}^n$ for most $d \in \mathbb{R}^n$. If $\beta \neq \emptyset$ and $f_i' \neq 0$ for every $i \in \beta$, then (87b) and (87c) fail almost surely. This is because, for any $d \in \mathbb{R}^n$, d_i is almost surely nonzero for $i = 1, 2, \ldots, n$, and, if $f_i' \neq 0$ for every $i \in \beta$, then $\langle f_i', d\rangle$ is almost surely nonzero for $i \in \beta$. If $\beta = \emptyset$, the conditions (87) can be simplified as follows. Solving (87a) for z_α and substituting z_α into (87d), we find that a solution of (87) requires some $z_\gamma \in \mathbb{R}^{|\gamma|}$ that solves

$$2\text{diag}\,(d_i/x_i^*)\langle f_i'(x^*), d\rangle + \langle (f_i')'(x^*; d), d\rangle - \langle f_{i,\alpha}'(x^*), z_\alpha\rangle \\ = \langle f_{i,\gamma}'(x^*), z_\gamma\rangle, \qquad \text{all } i \in \gamma. \quad (88)$$

Equation (88) is solvable only if the left-hand side, which is an element of $\mathbb{R}^{|\gamma|}$, lies in the subspace spanned by range $f_{\gamma,\gamma}'(x^*)$ as is required by the right-hand side. Since, by assumption, the (left) null space N_* is nontrivial, we have from (66) that $\ker(f_{\gamma,\gamma}'(x^*))^T$ is nontrivial. Hence, the complementary space range $f_{\gamma,\gamma}'(x^*)$ must be a strict subspace of $\mathbb{R}^{|\gamma|}$. It seems likely that this containment will typically fail for almost all directions $d \in \mathbb{R}^n$.

In summary, the starlike domain of convergence $\mathcal{R}$ is directionally dense at the solution x^* if *(1)* nonsingularity of (70) holds for almost every $d = t \in \mathcal{S}$, *(2)* for almost every $d \in \mathbb{R}^n$, the system of equations (87) fails to have a solution $z \in \mathbb{R}^n$, and *(3)* nonsingularity of (71) holds for almost every $d = g(t)/\|g(t)\|$ with $t \in \mathcal{S}$. Conditions *(1)* and *(2)* involve only the NCP function f, while condition *(3)* involves Ψ through the definition of g. If we assume that $N \cap \Pi_0^{-1}(0) = \{0\}$, then condition *(3)* is trivially satisfied because range $g = N$. The assumption $N \cap \Pi_0^{-1}(0) = \{0\}$ appears in Sect. 1 under the name $2^\forall$-regularity (Definition 4). As discussed in Sect. 1, $2^\forall$-regularity is a strong form of 2-regularity which, in particular, implies isolation of the solution. However, this assumption allows us to write the conditions ensuring directional denseness of the starlike domain of convergence entirely in terms of f, as we now state formally.

Theorem 6 *Consider a solution x^* of NCP(f) for $f : \mathbb{R}^n \to \mathbb{R}^n$ with f' strongly semismooth at x^*. Suppose that x^* is a singular solution in the sense that $N =$ ker $f'_{\gamma,\beta\cup\gamma}(x^*)$ is nontrivial. The starlike domain of convergence $\mathcal{R}$ for Newton's method (or $\mathcal{R}_\alpha$ for $\alpha \in [1, 2)$ for the 2-step accelerated Newton's method* (45) *and* (46)) *applied to the nonlinear-equations reformulation $\Psi(x)$ of NCP(f) is directionally dense if the following conditions hold:*

- (i) *the matrix* (70) *is nonsingular for almost every $d \in \mathbb{R}^n$,*
- (ii) *the system of equations* (87) *has no solution $z \in \mathbb{R}^n$ for almost every $d \in \mathbb{R}^n$, and*
- (iii) *the matrix* (71) *is nonsingular for every $d \in N \setminus \{0\}$.*

7 Numerical results on simple NCPs

We describe here some computational results obtained from a simple test set of NCPs of small dimension, defined in Appendix A. Properties of the problems are shown in Table 1. If the problem has more than one default starting point/solution pair, a numerical code is appended to the problem name. (These starting points and solutions are listed in Table 2.) The convergence rate shown in Table 1 is for Newton's method with unit step length. We also tabulate the sizes of the sets α, β, and γ, and the satisfaction of various rank and regularity properties at the solution in question. (2^T-regularity is defined in Definition 7, and 2^{ae}-regularity in Definition 5. For a definition of b-regularity, see [6, Definition 3.3.10].)

The solutions of our test problems are all isolated except for the solution $x^* = (0, 1)$ of the problems affknot1 and quadknot. 2^T-regularity fails at this solution for both of these problems, consistently with the fact that 2^T-regularity is sufficient for isolation. The 2^{ae}-regularity condition holds for quadknot at $x^* = (0, 1)$ and, as suggested by our theory, Newton's method converges from arbitrary, nearby starting points to this solution. For affknot1, 2^{ae}-regularity fails at $x^* = (0, 1)$, and we observe convergence to this solution only from points x_0 for which the projection of $x_0 - x^*$ onto the null space N (65) gives a direction for which 2-regularity holds. Specifically, for affknot1,

Table 1 Convergence rate of Newton's method on Ψ for the simple NCP test problems, showing regularity properties

Problem, s.p.	n	dim N	cgce rate	$\|\alpha\|$	$\|\beta\|$	$\|\gamma\|$	full rank		regularity		
							$f'_{\gamma,\gamma}$	$f'_{\gamma,\beta\cup\gamma}$	b	2^T	2^{ae}
quarp, 1	1	0	suplin	1	0	0	—	—	•	—	—
aff1	2	0	suplin	1	0	1	•	•	•	—	—
DIS61, 2	2	0	suplin	1	0	1	•	•	•	—	—
quarquad, 1	2	1	1/2	0	1	1	•	•	•	•	•
affknot1	2	1	1/2	0	1	1		•			
affknot2	2	1	1/2	0	1	1	•	•	•	•	•
quadknot	2	2	1/2	0	1	1					•
munson4	2	2	1/2	0	0	2				•	•
DIS61, 1	2	2	1/2	0	1	1				•	•
DIS64	2	2	1/2	0	2	0	—	—	•	•	•
ne-hard	3	2	1/2	0	2	1	•	•			•
doubleknot	4	2	1/2	0	2	2	•	•	•	•	•
quad1,1	2	1	1/2	0	1	1	•	•			
quad2,1	2	2	1/2	0	2	0					
quad1,2	2	1	2/3	0	1	1	•	•			
quad2,2	2	2	2/3	0	2	0					
quarquad, 2	2	1	3/4	1	0	1					
quarp, 2	1	1	3/4	0	0	1					
quarn	1	1	3/4	0	0	1					

• = property satisfied, blank = property not satisfied, — = property not applicable

we have $N = \{\delta e_2 \mid \delta \in \mathbb{R}\}$, and 2-regularity along $d = \delta e_2$ fails if $\delta \geq 0$ and holds if $\delta < 0$. Accordingly, Newton's method converges to $x^* = (0, 1)$ with rate $1/2$ from starting points $x_0 = (x_0^1, x_0^2)$ with $x_0^2 < 1$, while if $x_0^2 > 1$, Newton's method converges in one step to the solution $(0, x_0^2)$.

Only affknot1, quad1, and quad2 satisfy 2-regularity on a set of directions in N (or $\mathbb{R}^n$) having measure that is positive but less than 1. For affknot1, 2-regularity holds for half of the directions in N but almost every direction in $\mathbb{R}^n$. The problems quad1 and quad2 satisfy 2-regularity for half of the directions in both N and $\mathbb{R}^n$. As a result, convergence to the solutions of these problems occurs with two different rates. The first starting points for quad1 and quad2 in Table 1 demonstrate convergence along a direction satisfying 2-regularity with rate $1/2$, while the second starting points demonstrate convergence along a direction failing 2-regularity with slower convergence rate.

All problems but quarquad2, quarp,2, and quarn satisfy 2-regularity (67) for *some* $d \in \mathbb{R}^n$. Further, most of the problems also satisfy 2-regularity for almost every $d \in \mathbb{R}^n$; only the problems failing 2^{ae}-regularity, except for affknot1, fail to be 2-regular for almost every $d \in \mathbb{R}^n$.

In Table 3, we report the numbers of iterations required for local convergence of Newton's method and the Accelerated Newton method of Sect. 5 for the subset of

Table 2 Performance of accelerated Newton method (with $\alpha = 1.9$) for the NCP test problems for which the convergence rate of pure Newton is linear with factor $1/2$

problem, starting point	Newton iters	Accel Newton iters	Accel phase iters
quarquad,1	16	10	5
affknot1	20	10	7
affknot2	19	10	5
quadknot	18	8	5
munson4	19	12	4
DIS61, 1	19	12	5
DIS64	21	11	7
ne-hard	25	19	5
doubleknot	22	14	5
quad1, 1	15	9	4
quad2, 1	20	13	5

We show iterations for the pure Newton method, iterations for accelerated Newton method, and the iterations required by the accelerated Newton method in the accelerated phase, after a convergence rate of $1/2$ had been detected in the pure Newton method

Simple NCP test problems and starting points giving convergence rates for Newton's method of $1/2$. This is the subset of problems with a nontrivial null space N for which 2^{ae}-regularity may hold. (In fact, affknot1, quad1,1, and quad2,1 have convergence rates of $1/2$ for Newton's method but do not satisfy 2^{ae}-regularity. Despite the absence of 2^{ae}-regularity, the acceleration technique of Sect. 5 hastens the convergence.) We detect linear convergence at a rate of $1/2$ by applying the following tests to successive Newton steps p_i:

$$\left| \frac{\|p_i\|}{\|p_{i-1}\|} - \frac{\|p_{i-1}\|}{\|p_{i-2}\|} \right| < \texttt{cCauchy} \quad \text{and} \quad \left| \frac{\|p_i\|}{\|p_{i-1}\|} - \frac{1}{2} \right| < \texttt{cLinear}$$

with $\texttt{cCauchy} = .005$ and $\texttt{cLinear} = .01$. If both tests are satisfied at iteration i, we scale the next step p_{i+1} (and every second step thereafter) by the acceleration factor $\alpha = 1.9$. Convergence is declared when $\|\Psi(x)\| \leq 10^{-11}$.

The final column of Table 3 shows the number of steps taken in the "accelerated phase," following detection of a linear convergence rate in the pure Newton method. Note that the accelerated phase was present for all problem instances and that the number of steps taken in this phase is similar for all problems. For $\alpha = 1.9$, the convergence rate in the accelerated phase predicted by Theorem 2 was observed for all problems.

Appendix A: Simple NCP test set—problem descriptions

Below we list the simple NCP test problems, their solutions, and the corresponding starting points used to initialize Newton's method. A solution is any x satisfying

$$0 \leq x \;\perp\; f(x) \geq 0,$$

and we denote such x by x^*. Table 3 lists the starting point x_0 that was used for each solution x^*.

1. quarp

$$f(x) = (1 - x)^4$$

2. aff1

$$f(x) = \begin{bmatrix} x_1 + 2x_2 \\ x_2 - 1 \end{bmatrix}$$

3. DIS61 ([1, Example 6.1])

$$f(x) = \begin{bmatrix} (x_1 - 1)^2 \\ x_1 + x_2 + x_2^2 - 1 \end{bmatrix}$$

4. quarquad

$$f(x) = \begin{bmatrix} -(1 - x_1)^4 + x_2 \\ 1 - x_2^2 \end{bmatrix}$$

5. affknot1

$$f(x) = \begin{bmatrix} x_2 - 1 \\ x_1 \end{bmatrix}$$

6. affknot2

$$f(x) = \begin{bmatrix} x_2 - 1 \\ x_1 + x_2 - 1 \end{bmatrix}$$

7. quadknot

$$f(x) = \begin{bmatrix} x_2 - 1 \\ x_1^2 \end{bmatrix}$$

8. munson4 (from MCPLIB [17])

$$f(x) = \begin{bmatrix} -(x_2 - 1)^2 \\ -(x_1 - 1)^2 \end{bmatrix}$$

9. DIS64 ([1, Example 6.4])

$$f(x) = \begin{bmatrix} -x_1 + x_2 \\ -x_2 \end{bmatrix}$$

Table 3 Starting point/Solution pairs

Problem, s.p.	x_0	x^*
quarp, 1	0.1	0
aff1	(0.1, 0.9)	(0,1)
DIS61, 2	(0.2, 0.85)	$(0, (\sqrt{5}-1)/2)$
quarquad, 1	(0.1, 0.9)	(0, 1)
affknot1	(0.9, 0.1)	(0, 1)*
affknot2	(0.5, 0.5)	(0, 1)
quadknot	(0.5, 0.5)	(0, 1)*
munson4	(0, 0)	(1, 1)
DIS61, 1	(1.5, −0.5)	(1, 0)
DIS64	(2, 4)	(0, 0)
ne-hard	(10, 1, 10)	$(0, 0, \sqrt{200})$
doubleknot	(0.5, 0.5, 0.5, 0.5)	(1, 0, 0, 1)
quad1, 1	(0.9, −0.1)	(1, 0)
quad2, 1	(−1, −1)	(0, 0)
quad1, 2	(0.9, 0.1)	(1, 0)
quad2, 2	(1, 1)	(0, 0)
quarquad, 2	(0.9, 0.1)	(1, 0)
quarp, 2	0.9	1
quarn	0.9	1

*Full solution set is $(0, \delta)$ for $\delta \geq 1$

10. ne-hard (from MCPLIB [17])

$$f(x) = \begin{bmatrix} \sin x_1 + x_1^2 \\ x_2^3 + x_1 x_3 \\ x_3^2 - 200 + x_1 x_2 \end{bmatrix}$$

11. doubleknot

$$f(x) = \begin{bmatrix} 1 - x_1 + x_2 + x_3 \\ x_1 - 1 \\ x_4 - 1 \\ 1 + x_3 - x_4 \end{bmatrix}$$

12. quad1

$$f(x) = \begin{bmatrix} x_1 - 1 \\ x_2^2 \end{bmatrix}$$

13. quad2

$$f(x) = \begin{bmatrix} x_1^2 \\ x_2 \end{bmatrix}$$

14. quarn

$$f(x) = -(1 - x)^4$$

Acknowledgments We are grateful to two referees for their extremely helpful comments and suggestions.

References

1. Daryina, A.N., Izmailov, A.F., Solodov, M.V.: A class of active-set Newton methods for mixed complementarity problems. SIAM J. Optim. **15**(2), 409–429 (2004)
2. Decker, D.W., Keller, H.B., Kelley, C.T.: Convergence rates for Newton's method at singular points. SIAM J. Numer. Anal. **20**, 296–314 (1983)
3. Decker, D.W., Kelley, C.T.: Newton's method at singular points. I.. SIAM J. Numer. Anal. **17**(1), 66–70 (1980)
4. Decker, D.W., Kelley, C.T.: Convergence acceleration for Newton's method at singular points. SIAM J. Numer. Anal. **19**, 219–229 (1981)
5. Evtushenko, Y.G., Purtov, V.A.: Sufficient conditions for a minimum for nonlinear programming problems. Soviet Math. Dokl. **30**, 313–316 (1984)
6. Facchinei, F., Pang, J.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, Heidelberg (2003)
7. Golub, G.H., Van Loan, C.F.: Matrix Computations, third edn. The Johns Hopkins University Press, Baltimore (1996)
8. Griewank, A.: Starlike domains of convergence for Newton's method at singularities. Numer. Math. **35**(1), 95–111 (1980)
9. Griewank, A.: On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. **27**, 537–563 (1985)
10. Griewank, A., Osborne, M.R.: Newton's method for singular problems when the dimension of the null space is >1. SIAM J. Numer. Anal. **18**(1), 145–149 (1981)
11. Izmailov, A.F., Solodov, M.V.: Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. Math. Program. Ser. A **89**, 413–435 (2001)
12. Izmailov, A.F., Solodov, M.V.: Superlinearly convergent algorithms for solving singular equations and smooth reformulations of complementarity problems. SIAM J. Optim. **13**(2), 386–405 (2002)
13. Izmailov, A.F., Solodov, M.V.: The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. **27**, 614–635 (2002)
14. Josephy, N.H.: Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin-Madison (1979)
15. Kanzow, C.: Some equation-based methods for the nonlinear complementarity problem. Optim. Methods Softw. **3**(1), 327–340 (1994)
16. Kelley, C.T., Suresh, R.: A new acceleration method for Newton's method at singular points. SIAM J. Numer. Anal. **20**(5), 1001–1009 (1983)
17. ftp://ftp.cs.wisc.edu/math-prog/mcplib
18. Oberlin, C., Wright, S.J.: An accelerated Newton method for equations with equations with semismooth Jacobians and nonlinear complementarity problems: Extended version. Optimization technical report, Computer Sciences Department, University of Wisconsin-Madison (2006, Revised January 2007)
19. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
20. Reddien, G.W.: On Newton's method for singular problems. SIAM J. Numer. Anal. **15**(5), 993–996 (1978)
21. Reddien, G.W.: Newton's method and high order singularities. Comput. Math. Appl. **5**, 79–86 (1979)