Mathematical Programming manuscript No.
(will be inserted by the editor)

Christina Oberlin · Stephen J. Wright

An Accelerated Newton Method for Equations with Semismooth Jacobians and Nonlinear Complementarity Problems: Extended Version

Received: date / Accepted: date

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, under this regularity assumption, 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 formulation 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 formulation of the NCP at x^*.

Keywords Nonlinear Equations · Semismooth Functions · Newton's Method · Nonlinear Complementarity Problems

Mathematics Subject Classification (2000) 65H10 · 90C33

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.

Christina Oberlin
Computer Sciences Department,
University of Wisconsin,
1210 W. Dayton Street,
Madison, WI 53706. E-mail: coberlin@cs.wisc.edu

Stephen J. Wright
Computer Sciences Department,
University of Wisconsin,
1210 W. Dayton Street,
Madison, WI 53706. E-mail: swright@cs.wisc.edu

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 formulation of nonlinear complementarity problems (NCP) defined by

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

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), this nonlinear-equations formulation is not twice differentiable at x^*, and the weaker smoothness assumptions considered in this paper are required. Our results show that (i) the simple approach of applying Newton's method to the nonlinear-equations formulation of the NCP converges inside a starlike domain centered at x^*, albeit at a linear rate if the solution is singular; (ii) a simple technique can be applied to accelerate the convergence in this case 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] and the discussion in Facchinei and Pang [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 [19], 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 partial neighborhood of x^* known as a *starlike domain*, can be extended to the case in which F' is strongly semismooth at x^*; see Section 4. In Section 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 Section 6, we consider a particular nonlinear-equations reformulation of the NCP and interpret the regularity conditions for this formulation as conditions on the NCP. We show that they reduce to previously known NCP regularity conditions in certain special cases. We conclude in Section 7 by presenting computational results for some simple NCPs, including a number of degenerate examples.

We start with certain preliminaries and definitions of notation and terminology (Section 2), followed by a discussion of prior relevant work (Section 3).

2 Definitions and Properties

For $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$ we denote the derivative by $G' : \Omega \to \mathbb{R}^{p\times n}$, that is,

$$G'(x) = \begin{bmatrix} \frac{\partial G_1}{\partial x_1} & \cdots & \frac{\partial G_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial G_p}{\partial x_1} & \cdots & \frac{\partial G_p}{\partial x_n} \end{bmatrix}. \tag{2}$$

For a scalar function $g : \Omega \to \mathbb{R}$, the derivative $g' : \Omega \to \mathbb{R}^n$ is the vector function

$$g'(x) = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{bmatrix}.$$

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 $\operatorname{range} M$. $\operatorname{rank} M$ denotes the rank of the matrix M, which is the dimension of $\operatorname{range} M$.

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 $\lambda > 0$.

2.1 Smoothness Conditions

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

$$\lim_{t\downarrow 0} \frac{G(x+td) - G(x)}{t} \tag{3}$$

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

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. We say that 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, if G is strongly semismooth at every $\bar{x} \in \Omega$, we say that G is strongly semismooth on Ω.

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 [18]. Hence, for $F' : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ strongly semismooth at x^*, there is some L_{x^*} such that

$$\|(F')'(x^*;h_1) - (F')'(x^*;h_2)\| \leq L_{x^*}\|h_1 - h_2\|. \tag{4}$$

Provided 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].

$$F'(x) = F'(x^*) + (F')'(x^*; x - x^*) + O(\|x - x^*\|^2). \tag{5}$$

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

2.2 2-regularity

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 complement of N, such that $N \oplus N_\perp = \mathbb{R}^n$, and let $N_* := \ker F'(x^*)^T$ with 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

$$(P_{N_*}F')'(x^*;d)|_N \quad \text{is nonsingular.} \tag{6}$$

The 2-regularity conditions of Reddien [19], Decker and Kelley [3], and Griewank [8] require (6) to hold for certain $d \in N$. In fact, this property first appeared in the literature as $(P_{N_*}F''(x^*)d)|_N$; the form in (6) was introduced by Izmailov and Solodov in [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 [13].

Decker and Kelley [3] and Reddien [19] 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 (6) holds for every $d \in N \setminus \{0\}$.

For F twice differentiable at x^*, $2^\forall$-regularity implies (geometric) isolation of the solution x^* [19,3] 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 (6) holds for almost every $d \in N$.

The following example due to Griewank [9, p. 542] shows that a 2^{ae}-regular solution need not be isolated, when $\dim N > 1$. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as

$$F(x_1, x_2) = \begin{bmatrix} x_1^2 \\ x_1x_2 \end{bmatrix}, \tag{7}$$

with roots $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0\}$. It can be verified that the origin is a 2^{ae}-regular root of this function. First note that $N \equiv \mathbb{R}^2 \equiv N_*$. By Definition 5, 2^{ae}-regularity holds if $F''(x^*)d$ is nonsingular for almost every $d = (d_1, d_2)^T \in N$. By direct calculation, we have

$$F''(x^*)d = \begin{bmatrix} 2d_1 & 0 \\ d_2 & d_1 \end{bmatrix},$$

which is nonsingular whenever $d_1 = 0$, that is, for almost every $d \in \mathbb{R}^2$.

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

Definition 6 $\mathbf{2^1}$-regularity. 2^1*-regularity* holds for F at x^* if (6) 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 (9) of the nonlinear complementarity problems quad2 and affknot1 (defined in Appendix C) are 2^1-regular but not 2^{ae}-regular at their solutions.

Izmailov and Solodov introduce a regularity condition and prove that it implies that x^* is an isolated solution, provided that $P_{N_*}F'(x^*)$ is B-differentiable [11, Theorem 5(a)]. 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

$$T_2 := \{d \in N \mid (P_{N_*}F')'(x^*; d)d = 0\}. \tag{8}$$

Definition 7 $\mathbf{2^T}$-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 Section 7, neither 2^T-regularity nor 2^{ae}-regularity implies the other. 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). For completeness, we verify this claim. Suppose $N =$ span v, for $v \in \mathcal{S}$. By positive homogeneity of the directional derivative, 2^T-regularity holds if $(P_{N_*}F')'(x^*; v)v \neq 0$ and $(P_{N_*}F')'(x^*; -v)(-v) \neq 0$. Similarly, the definition of $2^\forall$-regularity requires $(P_{N_*}F')'(x^*; v)|_N$ and $(P_{N_*}F')'(x^*; -v)|_N$ to be nonsingular. By linearity, we need to verify only that $(P_{N_*}F')'(x^*; v)v \neq 0$ and $(P_{N_*}F')'(x^*; -v)(-v) \neq 0$, equivalently, that 2^T-regularity is satisfied.

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.

3 Prior Work

In this section, we summarize briefly the prior work most 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 on the use of 2-regularity conditions to prove convergence of Newton-like methods to singular solutions. As explained in Subsection 2.2, such conditions concern the behavior of the directional derivative of F' on the null spaces N and N_* of $F'(x^*)$ and $F'(x^*)^T$, respectively.

The 2^1-regularity condition (Definition 6) was used in [20] by Reddien and in [10] by Griewank and Osborne. 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 (6) 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 [19,3,2] is much simpler than the analysis presented in 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^*, then the starlike domain is nonempty. As mentioned in Subsection 2.2, 2^1-regularity is equivalent to 2^{ae}-regularity when F is twice Lipschitz continuously differentiable at x^*. In this case, 2^{ae}-regularity implies that 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 an acceleration 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] prove superlinear convergence of an accelerated scheme under less stringent assumptions. If 2^1-regularity holds at x^* and the third derivative of F at x^* is bounded over the truncated cone about N, then overrelaxing every other step by a factor approaching 2 results in superlinear convergence.

By contrast, the acceleration technique that we analyze in Section 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 use a scheme like that of Kelley and Suresh [16], increasing the overrelaxation factor to achieve superlinear convergence.

Smooth Nonlinear-Equations Formulation of the NCP. In the latter part of this paper, we discuss a nonlinear-equations formulation 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

$$\Psi_i(x) := 2x_i f_i(x) - (\min(0, x_i + f_i(x)))^2, \quad i = 1, 2, \ldots, n. \tag{9}$$

This formulation 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^*. At a solution x^* for which $x_i^* = f_i(x^*) = 0$ for some i, x^* is a singular root of Ψ and Ψ 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.) In particular, 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 $f'(x^*)$. These conditions are weaker than those required for superlinear convergence of known nonsmooth-nonlinear-equations formulations of NCP. In [12], Izmailov and Solodov augment the formulation $\Psi(x) = 0$ 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. The quasi-regularity condition resembles the 2-regularity condition for the NCP; their relationship is discussed in Subsection 6.3 below.

In contrast to the algorithms of [1] and [12], the approach we present in this work 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 C 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 [19,3] 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 that 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 certain vector s in the null space N. The domain $\mathcal{W}_s$ is similar to the domains of convergence found in earlier works (Reddien [19], Decker and Kelley [3]). 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 $\mathcal{R} - x^*$ has zero measure. As a result, the error in the initial iterate with respect to x^* need not lie near the null space N [8].

In this section, we weaken the smoothness assumption of Griewank in replacing the second derivative of F in (6) 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 (10) 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 (32).*

Only a few modifications to Griewank's proof [8] are necessary. We use the properties (4) and (5) to show that F is smooth enough for the main 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.

For completeness, we work through the details of the proof in the remainder of this section and in Section A in the appendix, highlighting the points of departure from Griewank's proof as they arise.

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

$$x^* = 0, \quad F'(x^*) = F'(0) = (I - P_{N_*}), \text{ and } N_* = \mathbf{R}^m \times \{0\}^{n-m}. \tag{10}$$

(We revoke assumption (10) in our discussion of an equation reformulation of the NCP in Sections 6 and 7.)

For $x \in \mathbf{R}^n \setminus \{0\}$, 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 (10), 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 (10), we obtain

$$F'(0) = \begin{bmatrix} 0_{m\times m} & 0_{m\times n-m} \\ 0_{n-m\times m} & I_{n-m\times n-m} \end{bmatrix}. \tag{11}$$

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

Using (10), 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 (11), 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

$$\bar{B}(x) = \bar{B}(x - x^*) = (P_{N_*}F')'(x^*; x - x^*)|_N = (P_{N_*}F')'(0; x)|_N, \tag{12a}$$

$$\bar{C}(x) = \bar{C}(x - x^*) = (P_{N_*}F')'(x^*; x - x^*)|_{N_\perp} = (P_{N_*}F')'(0; x)|_{N_\perp}. \tag{12b}$$

From $x = \rho t$, the expansion (5) with $x^* = 0$ yields

$$\begin{aligned} B(x) &= \bar{B}(x) + O(\rho^2) = \rho\bar{B}(t) + O(\rho^2), \\ C(x) &= \bar{C}(x) + O(\rho^2) = \rho\bar{C}(t) + O(\rho^2), \\ D(x) &= O(\rho), \quad \text{and} \quad E(x) = I + O(\rho). \end{aligned} \tag{13}$$

Note that 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 (5) to arrive at (13), 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 (13), we have

$$G(x) = B(x) + O(\rho^2) = \rho\bar{B}(t) + O(\rho^2). \tag{14}$$

Hence,

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

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 by the expression above that

$$\nu(\rho t) = O((\det F'(\rho t))^{1/m}) = \begin{cases} O(\rho), & \text{if } \bar{B}(t) \text{ is nonsingular,} \\ o(\rho), & \text{if } \bar{B}(t) \text{ is singular.} \end{cases} \tag{15}$$

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

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

Following Griewank [8], we define the function $\sigma(t)$ to be the minimum of 1 and the L_2 operator norm of the smallest singular value of $\bar{B}(t)$, that is,

$$\sigma(t) := \begin{cases} 0 & \text{if } \bar{B}(t) \text{ is singular} \\ \min(1, \|\bar{B}^{-1}(t)\|^{-1}) & \text{otherwise.} \end{cases} \tag{16}$$

It is a fact from linear algebra that the individual singular values of a matrix vary continuously with respect to perturbations of the matrix [7, Theorem 8.6.4]. By (4), $\bar{B}(t)$ 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 (4) 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 stronger smoothness assumptions.

Let

$$\Pi_0(d) := \det\bar{B}(d), \qquad \text{for } d \in \mathbb{R}^n. \tag{17}$$

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 necessarily 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 $\mathcal{R}$. The definitions of $\mathcal{W}_s$ and $\mathcal{R}$ depend on several functions that we first introduce. If we define $\min(\emptyset) = \pi$, the angle

$$\phi(s) := \frac{1}{4} \min\{\cos^{-1}(t^T s) \mid t \in \mathcal{S} \cap {\Pi_0}^{-1}(0)\}, \quad \text{for } s \in N \cap \mathcal{S} \tag{18}$$

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 assertion 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 analysis. 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 (4).

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

$$\bar{\mathcal{R}} := \{x = \rho t \mid t \in \mathcal{S},\, 0 < \rho < \bar{r}(t)\}, \tag{19}$$

where

$$\bar{r}(t) := \min\left\{r_b, \frac{1}{2}\gamma^{-1}\sigma(t)\right\}. \tag{20}$$

Note that the excluded directions, $t \in \mathcal{S}$ 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 (15) and (16). Even if $\sigma(t) \neq 0$ for some $t \in \mathcal{S}$, the set of excluded directions may have positive measure in $\mathcal{S}$. This is the case for the the smooth nonlinear equations reformulation (9) of the nonlinear complementarity problems quad2 (defined in Appendix C). 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

$$\hat{r}(s) := \min\{\bar{r}(t) \mid t \in \mathcal{S},\, \cos^{-1}(t^T s) \leq \phi(s)\}, \text{ for } s \in N \cap \mathcal{S} \tag{21}$$

and

$$\hat{\sigma}(s) := \min\{\sigma(t) \mid t \in \mathcal{S},\, \cos^{-1}(t^T s) \leq \phi(s)\}, \text{ for } s \in N \cap \mathcal{S}. \tag{22}$$

These minima exist and both are nonnegative and continuous on $\mathcal{S} \cap N$ with $\hat{\sigma}^{-1}(0) = \hat{r}^{-1}(0) = \phi^{-1}(0)$. Note that 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

$$c := \max\{\|\bar{C}(t)\| + \sigma(t) \mid t \in \mathcal{S}\}. \tag{23}$$

In the following, we use the abbreviation

$$q(s) := \frac{1}{4}\sin\phi(s) \le \frac{1}{4}, \text{ for } s \in N \cap \mathcal{S}. \tag{24}$$

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

$$\sin\hat{\phi}(s) := \min\left\{ \frac{q(s)}{c/\hat{\sigma}(s) + 1 - q(s)}, \frac{2\delta\hat{r}(s)}{(1-q(s))\hat{\sigma}^2(s)} \right\}, \text{ for } s \in N \cap \mathcal{S}, \tag{25}$$

where δ is a problem-dependent, positive number to be specified in the Appendix in (144). We now define

$$\hat{\rho}(s) := \frac{(1-q(s))\hat{\sigma}^2(s)}{2\delta}\sin\hat{\phi}(s), \text{ for } s \in N \cap \mathcal{S}. \tag{26}$$

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

$$\hat{\phi}^{-1}(0) = \hat{\rho}^{-1}(0) = \phi^{-1}(0) = \Pi_0^{-1}(0) \cap N \cap \mathcal{S}. \tag{27}$$

Now we can define the starlike domain $\mathcal{W}_s$,

$$\mathcal{W}_s := \{x = \rho t \,|\, t \in \mathcal{S}, \cos^{-1}(t^T s) < \hat{\phi}(s), 0 < \rho < \hat{\rho}(s)\}, \tag{28}$$

and the starlike domain $\mathcal{I}_s$,

$$\mathcal{I}_s := \{x = \rho t \mid t \in \mathcal{S}, \cos^{-1}(t^T s) < \phi(s), 0 < \rho < \hat{\rho}(s)\}. \tag{29}$$

By the first inequality in (25), $\sin\hat{\phi}(s) \le \sin\phi(s)$. Since both $\hat{\phi}(s)$ and $\phi(s)$ are acute angles, we have $\hat{\phi}(s) \le \phi(s)$ 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

$$\hat{\rho}(s) \le \hat{r}(s) \le \bar{r}(t) \le r_b, \text{ for all } t \in \mathcal{S} \text{ with } \cos^{-1} t^T s \le \phi(s). \tag{30}$$

It follows that

$$\mathcal{I}_s \subset \bar{\mathcal{R}}, \text{ for all } s \in \mathcal{S} \cap N \setminus \Pi_0^{-1}(0). \tag{31}$$

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

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

$$g(x) = \rho g(t) = \begin{bmatrix} I & \bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x. \tag{32}$$

It is shown in (145) of Appendix A that the Newton iteration from a point x near x^* is, to first order, the map $x^* + \frac{1}{2}g(x)$, 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):

$$\mathcal{R} := \{x = \rho t \,|\, t \in \mathcal{S}, 0 < \rho < r(t)\}, \tag{33}$$

where

$$r(t) := \min\left\{\bar{r}(t), \frac{\sigma^2(t)\hat{\rho}(s(t))}{2\delta r_b + c\sigma(t) + \sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t)\sin\hat{\phi}(s(t))}{2\delta}\right\}, \tag{34}$$

where we define

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

and δ is the constant to be defined in (144) in the Appendix. (The factor of 2, or $k+1$ for the general case, is missing from the denominator of the second term in the definition of $r(t)$ in [8] but should have been included, as it is necessary for the proof of convergence.)

The remaining details of the proof of Theorem 1 appear in an appendix (Appendix A), which picks up the development at this point.

The Excluded Directions of $\mathcal{R}$. We conclude this section by characterizing the excluded directions of $\mathcal{R}$, that is, $t \in \mathcal{S}$ for which $r(t) = 0$. By the definition of $r(t)$ (34), 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. Let us inspect each of these possibilities. By definition, $\bar{r}(t)$ (20) 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 (27) $\hat{\rho}(s(t))$ 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:

$$t \in \Pi_0^{-1}(0), \quad g(t) = 0, \quad \text{or} \quad g(t)/\|g(t)\| \in \Pi_0^{-1}(0). \tag{35}$$

The first condition fails if F satisfies 2-regularity (6) for t. Likewise, the third condition fails if F satisfies 2-regularity (6) for $g(t)/\|g(t)\|$. Let us consider the second condition. For $d \in \mathbf{R}^n \setminus \Pi_0^{-1}(0)$, by the definition of g (32) we have

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

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 (12a) and (12b), we have

$$g(d) = 0 \Leftrightarrow (P_{N_*}F')'(x^*; d)d = 0, \qquad \text{for } d \in \mathbf{R}^n \setminus \Pi_0^{-1}(0). \tag{36}$$

The right-hand side of this condition is identical to the condition defining the set T_2 (8), 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 $\mathbf{R}^n$ to 0. This is despite the facts that Π_0 g, and $\Pi_0(g(\cdot))$ may be nonzero elsewhere, Π_0 and g are continuous and positively homogeneous, and g is the identity on its range N. Hence, the set of excluded directions can be of positive measure.

5 Acceleration of Newton's Method

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

$$x_{j+1} = x_j - \alpha F'(x_j)^{-1} F(x_j), \tag{37}$$

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

If every step is overrelaxed, it can be shown that the condition $\alpha < \frac{4}{3}$ must be satisfied to ensure convergence and, as a result, the rate of linear convergence is no faster than $\frac{1}{3}$.

In this section, 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 (37) for some fixed $\alpha \in [1, 2)$. Broadly speaking, each pure Newton step refocuses the error along the null space N. Kelley and Suresh prove superlinear convergence for this method when α is systematically increased to 2 as the iterates converge [16]. However, their proof requires the third derivative of F evaluated at x^* to satisfy a boundedness condition and assumes a starting point x_0 that lies near a 2^1-regular direction in N.

We state our main result here and prove it in the remainder of this section. The major assumptions are that 2^1-regularity holds at x^* and that $x_0 \in \mathcal{R}_\alpha$, where $\mathcal{R}_\alpha$ is a starlike domain defined in (50) whose excluded directions are identical to those of $\mathcal{R}$ defined in Section 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 for $j = 0, 1, 2, \ldots$ we have*

$$x_{2j+1} = x_{2j} - F'(x_{2j})^{-1} F(x_{2j}) \quad \text{and} \tag{38}$$

$$x_{2j+2} = x_{2j+1} - \alpha F'(x_{2j+1})^{-1} F(x_{2j+1}), \tag{39}$$

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

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

The remainder of this section contains the proof of the theorem.

5.1 Definitions

We assume the problem is in standard form (10). We define the positive constant $\tilde{\delta}$ as follows:

$$\tilde{\delta} := \delta \max(c, \alpha), \tag{40}$$

where δ is defined in (144) and c is defined in (23). Note that

$$\tilde{\delta} \geq \delta. \tag{41}$$

We introduce the following new parameters:

$$q_\alpha(s) := \frac{1-\alpha/2}{4}\sin\phi(s), \text{ for } s \in N \cap \mathcal{S}, \tag{42}$$

(from which it follows immediately that $q_\alpha(s) \le (1/8)\sin\phi(s) \le 1/8$). We define the angle $\tilde\phi_\alpha(s)$, for which $0 \le \tilde\phi_\alpha(s) \le \pi/2$, by the equality

$$\sin\tilde\phi_\alpha(s) := \min\left\{\frac{q_\alpha(s)}{c/\hat\sigma(s)+1-q_\alpha(s)}, \frac{2\delta\hat r(s)}{(1-q_\alpha(s))\hat\sigma^2(s)}\right\}, \text{ for } s \in N \cap \mathcal{S}. \tag{43}$$

Since $\alpha \ge 1$, a comparison of (24) and (42) yields

$$q_\alpha(s) \le \frac{1}{2}q(s). \tag{44}$$

The definition of $\sin\tilde\phi_\alpha$ (43) is simply that of $\sin\hat\phi$ (25) with q replaced by q_α. By (44), the numerators in the definition of $\sin\tilde\phi_\alpha$ are smaller or the same as those in the definition of $\sin\hat\phi$ and the denominators are larger or the same. As a result, we have

$$\sin\tilde\phi_\alpha(s) \le \sin\hat\phi(s) \le \sin\phi(s), \tag{45}$$

and therefore

$$\tilde\phi_\alpha(s) \le \phi(s). \tag{46}$$

We further define

$$\tilde\rho_\alpha(s) := \frac{(1-\alpha/2-q_\alpha(s))\hat\sigma^3(s)}{4\tilde\delta}\sin\tilde\phi_\alpha(s) \text{ for } s \in N \cap \mathcal{S}, \tag{47}$$

$$\mathcal{W}_{s,\alpha} := \{x = \rho t \,|\, t \in \mathcal{S},\ \cos^{-1}(t^T s) < \tilde\phi_\alpha(s), 0 < \rho < \tilde\rho_\alpha(s)\}, \tag{48}$$

and

$$\mathcal{I}_{s,\alpha} := \{x = \rho t \,|\, t \in \mathcal{S},\ \cos^{-1}(t^T s) < \phi(s), 0 < \rho < \tilde\rho_\alpha(s)\}. \tag{49}$$

(Note that (46) implies that $\mathcal{W}_{s,\alpha} \subseteq \mathcal{I}_{s,\alpha}$.) We will show that the following set is a starlike domain of convergence:

$$\mathcal{R}_\alpha := \{x = \rho t \,|\, t \in \mathcal{S},\ 0 < \rho < r_\alpha(t)\}, \tag{50}$$

where

$$r_\alpha(t) := \min\left\{\bar r(t), \frac{\sigma^2(t)\tilde\rho_\alpha(s(t))}{2\delta r_b + c\sigma(t) + \sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t)(1-\alpha/2)\sin\tilde\phi_\alpha(s(t))}{8\delta}\right\} \tag{51}$$

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

We now establish that $\mathcal{R}_\alpha \subseteq \mathcal{R} \subseteq \bar{\mathcal{R}}$ and $\mathcal{I}_{s,\alpha} \subseteq \mathcal{I}_s \subseteq \bar{\mathcal{R}}$. We first show that

$$\tilde{\rho}_\alpha(s) \le \hat{\rho}(s), \tag{52}$$

where $\hat{\rho}(s)$ is defined in (26). Because of (45), it suffices for (52) to prove that

$$\frac{(1-\alpha/2-q_\alpha(s))\hat{\sigma}^3}{4\tilde{\delta}} \le \frac{1-q(s)}{2\delta}\hat{\sigma}^2.$$

The truth of this inequality follows from $\alpha \in [1,2)$, $q(s) \in [0,\frac{1}{4}]$, $q_\alpha(s) \in [0,\frac{1}{4}-\frac{\alpha}{8}]$, $\hat{\sigma} \le 1$, and (41). Using (45) and (52) together with (41), a comparison of $r_\alpha(t)$ (51) and $r(t)$ (34) yields $r_\alpha(t) \le r(t)$. We conclude by comparing (33) with (50) that $\mathcal{R}_\alpha \subseteq \mathcal{R}$. The relation $\mathcal{R} \subseteq \bar{\mathcal{R}}$ follows easily from the definitions of r (34) and $\bar{r}$ (20) together with (33) and (19). By (52), we also have $\mathcal{I}_{s,\alpha} \subseteq \mathcal{I}_s$, upon comparing their definitions (49) and (29). The relation $\mathcal{I}_s \subseteq \bar{\mathcal{R}}$ was demonstrated in (31).

As in Section 4, we denote the sequence of iterates by $\{x_i\}_{i\ge 0}$ and use the notation (146), that is,

$$\rho_i = \|x_i\|, \quad t_i = x_i/\rho_i, \quad \sigma_i = \sigma(t_i), \quad s_i = g(x_i)/\|g(x_i)\|,$$

where $g(\cdot)$ is defined in (32). We use the following abbreviations throughout the remainder of this section:

$$\tilde{\rho}_\alpha \equiv \tilde{\rho}_\alpha(s_0), \quad \tilde{\phi}_\alpha \equiv \tilde{\phi}_\alpha(s_0), \quad \phi \equiv \phi(s_0), \quad \hat{\sigma} \equiv \hat{\sigma}(s_0). \tag{53}$$

5.2 Basic Error Bounds and Outline of Proof

Since the problem is in standard form, we have from (145) that the Newton step (38) satisfies the following relationships for $x_{2k} \in \bar{\mathcal{R}}$:

$$x_{2k+1} = \frac{1}{2}\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + e(x_{2k}) = \frac{1}{2}g(x_{2k}) + e(x_{2k}), \tag{54}$$

for all $k \ge 0$, where $g(\cdot)$ is defined in (32) and the remainder term $e(\cdot)$ is defined in (143). As in (144), we have

$$\|e(x_{2k})\| \le \delta\frac{\rho_{2k}^2}{\sigma_{2k}^2}. \tag{55}$$

For the accelerated Newton step (39), we have for $x_{2k+1} \in \bar{\mathcal{R}}$ that

$$x_{2k+2} = \begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix} x_{2k+1} + \alpha e(x_{2k+1}), \tag{56}$$

for all $k \ge 0$, which from (144) yields

$$\begin{aligned} \|x_{2k+2} - (1-\alpha/2)x_{2k+1}\| &\le \left\| \begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & -\frac{\alpha}{2}I \end{bmatrix} t_{2k+1} \right\| \rho_{2k+1} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\le \left\| \begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & -\frac{\alpha}{2}I \end{bmatrix} t_{2k+1} \right\| \rho_{2k+1} + \tilde{\delta}\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2}, \end{aligned} \tag{57}$$

where $\tilde{\delta}$ is defined in (40).

By substituting (54) into (56), we obtain

$$x_{2k+2} = \frac{1}{2}\begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix}\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + \tilde{e}_\alpha(x_{2k}, x_{2k+1}), \tag{58}$$

where

$$\tilde{e}_\alpha(x_{2k}, x_{2k+1}) = \begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix} e(x_{2k}) + \alpha e(x_{2k+1}). \tag{59}$$

Therefore,

$$\begin{aligned} x_{2k+2} &= \frac{1}{2}\left(1-\frac{\alpha}{2}\right)\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + \tilde{e}_\alpha(x_{2k}, x_{2k+1}) \\ &= \frac{1}{2}\left(1-\frac{\alpha}{2}\right) g(x_{2k}) + \tilde{e}_\alpha(x_{2k}, x_{2k+1}), \end{aligned} \tag{60}$$

To bound the remainder term, note that $|1-\frac{\alpha}{2}| + |1-\alpha| = \frac{\alpha}{2}$ for $\alpha \in [1, 2)$. Hence, we have from (59) that

$$\begin{aligned} \|\tilde{e}_\alpha(x_{2k}, x_{2k+1})\| &\le \frac{\alpha}{2}\left(1 + \|\bar{B}(t_{2k+1})^{-1}\|\|\bar{C}(t_{2k+1})\|\right)\|e(x_{2k})\| + \alpha\|e(x_{2k+1})\| \\ &\le \left(\frac{\sigma_{2k+1} + \|\bar{C}(t_{2k+1})\|}{\sigma_{2k+1}}\right)\delta\frac{\rho_{2k}^2}{\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\qquad \text{from } \alpha < 2, (16), \text{ and } (144) \\ &\le c\delta\frac{\rho_{2k}^2}{\sigma_{2k+1}\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\qquad \text{from } (23) \\ &\le \tilde{\delta}\frac{\rho_{2k}^2 + \rho_{2k+1}^2}{\mu_{2k}^3}, \end{aligned} \tag{61}$$

where

$$\mu_{2k} := \min(\sigma_{2k}, \sigma_{2k+1}) \tag{62}$$

and $\tilde{\delta}$ is defined as in (40). By combining (60) with (61), we obtain

$$\left\| x_{2k+2} - \frac{1}{2}\left(1-\frac{\alpha}{2}\right) g(x_{2k}) \right\| \le \tilde{\delta}\frac{\rho_{2k}^2 + \rho_{2k+1}^2}{\mu_{2k}^3}. \tag{63}$$

In other words, if $x_{2k} = \rho_{2k}t_{2k}$ with $t_{2k} \in \mathcal{S}$ and $x_{2k+1} = \rho_{2k+1}t_{2k+1}$ with $t_{2k+1} \in \mathcal{S}$ are sufficiently close to x^* and $\sigma(t_{2k})$ and $\sigma(t_{2k+1})$ are bounded below by a positive number, then the accelerated Newton iterate 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, $\frac{1}{2}(1-\frac{\alpha}{2})g(x_{2k})$ is a first order approximation to the double step achieved by applying a Newton step followed by an accelerated Newton step from x_{2k}.

Before proceeding with the proof, we state the definitions of certain quantities that appear in the appendix. The angle between iterate x_i and the null space N is denoted by θ_i, while ψ_i denotes the angle between x_i and s_0 (147). The proof of Theorem 2 is by induction. The induction step consists of showing that if

$$\rho_{2k+\iota} < \tilde{\rho}_\alpha, \quad \theta_{2k+\iota} < \tilde{\phi}_\alpha, \text{ and } \psi_{2k+\iota} < \phi, \tag{64}$$
$$\text{for } \iota \in \{1,2\}, \quad \text{all } k \text{ with } 0 \le k < j,$$

then

$$\rho_{2j+\iota} < \tilde{\rho}_\alpha, \quad \theta_{2j+\iota} < \tilde{\phi}_\alpha, \text{ and } \psi_{2j+\iota} < \phi \quad \text{for } \iota \in \{1,2\}, \tag{65}$$

where $\tilde{\rho}_\alpha$, $\tilde{\phi}_\alpha$, and ϕ are defined in (53). For all $i = 1, 2, \ldots$, the third property in (64) and (65)—$\psi_i < \phi$—implies the crucial fact that $\sigma_i \ge \hat{\sigma} > 0$; see (22) and (157). 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 Subsection A.1 together with (54) and (56) are valid for our iterates. The convergence rate claimed in the theorem is a byproduct of the proof of the induction step.

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 (64) for $j = 1$, as we now verify. By the definition of (48) (with $s := s_0$), $x_1 \in \mathcal{W}_{s_0,\alpha}$ implies that $\rho_1 < \tilde{\rho}_\alpha$ and $\psi_1 < \tilde{\phi}_\alpha$. Because of (46) and the elementary inequality $\theta_1 \le \psi_1$, we have $\theta_1 \le \psi_1 < \tilde{\phi}_\alpha \le \phi$. Therefore, the inequalities in (64) hold for $k = 0$ and $\iota = 1$. Since $x_2 \in \mathcal{I}_{s_0,\alpha}$, we have from (49) that $\rho_2 < \tilde{\rho}_\alpha$ and $\psi_2 < \phi$. With the additional fact that $\theta_2 < \tilde{\phi}_\alpha$, we conclude that the inequalities in (64) hold for $k = 0$ and $\iota = 2$. Hence, (64) holds for $j = 1$.

5.3 The Anchor Step

We begin by proving the anchor step. The proof of Theorem 1 shows that if $x_0 \in \mathcal{R}$ then $x_1 \in \mathcal{W}_{s_0}$. We show in a similar fashion that if $x_0 \in \mathcal{R}_\alpha$ then $x_1 \in \mathcal{W}_{s_0,\alpha}$. Since the first step is a Newton step from $x_0 \in \mathcal{R}_\alpha$ and since $\mathcal{R}_\alpha \subseteq \mathcal{R}$, the inequalities of Section 4 and Appendix A remain valid. In particular, we can reuse (155) and write

$$\rho_1 \le \rho_0 \left(\frac{1}{2}\left(1 + \frac{c}{\sigma_0}\right) + \delta \frac{\rho_0}{\sigma_0^2} \right) = \frac{1}{2}\rho_0 \frac{\sigma_0^2 + c\sigma_0 + 2\delta\rho_0}{\sigma_0^2}. \tag{66}$$

Since $x_0 \in \mathcal{R}_\alpha$, we have $\rho_0 < r_\alpha(t_0)$ by (50). In addition, since $r_\alpha(t_0) \le \bar{r}(t_0) \le r_b$ (which follows from (20) and (51)), we have $\rho_0 < r_b$. Hence, from (66), we have

$$\rho_1 < \frac{1}{2} r_\alpha(t_0) \frac{\sigma_0^2 + c\sigma_0 + 2\delta r_b}{\sigma_0^2}.$$

By using the second part of the definition of r_α (51), we thus obtain

$$\rho_1 < \frac{1}{2}\tilde{\rho}_\alpha < \tilde{\rho}_\alpha. \tag{67}$$

As noted above, the inclusion $\mathcal{R}_\alpha \subseteq \mathcal{R}$ implies that inequality (150) is valid here, that is,

$$\sin\psi_1(s_0) \le \left(\frac{1}{2}\|g(t_0)\|\right)^{-1} \delta\frac{\rho_0}{\sigma_0^2}.$$

Since $\rho_0 < r_\alpha(t_0)$, we can apply the third inequality implicit in the definition of r_α (51) to obtain

$$\sin\psi_1(s_0) \le \frac{1-\alpha/2}{4}\sin\tilde{\phi}_\alpha(s_0) < \sin\tilde{\phi}_\alpha(s_0). \tag{68}$$

It is shown in Appendix B that $\psi_1 \le \pi/2$, and thus $\psi_1(s_0) < \tilde{\phi}_\alpha(s_0)$. The bounds (67) and (68) together show that $x_1 \in \mathcal{W}_{s_0,\alpha}$. We note that (68) and (46) imply that $\psi_1 < \tilde{\phi}_\alpha \le \phi$, which implies $\sigma_1 \ge \hat{\sigma}(s_0) = \hat{\sigma}$, by the definition of $\hat{\sigma}$ (22).

Next we show that if $x_0 \in \mathcal{R}_\alpha$, then $x_2 \in \mathcal{I}_{s_0,\alpha}$. We begin by showing that $\rho_2 < \rho_1$, from which $\rho_2 < \tilde{\rho}_\alpha$ follows from (67). From (56) for $k = 0$, we have by decomposing x_1 into components in N and $N_\perp$ that

$$\begin{aligned}
\rho_2 &\le \left(\left(1-\frac{\alpha}{2}\right)\cos\theta_1 + \left(\frac{\alpha\|\bar{C}(t_1)\|}{2\sigma_1} + \alpha - 1\right)\sin\theta_1 + \alpha\delta\frac{\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\qquad \text{from (16), (55), and (144) with } x = x_1 \\
&\le \left(1-\frac{\alpha}{2} + \left(\frac{\alpha}{2}\frac{\|\bar{C}(t_1)\|+\sigma_1}{\sigma_1} + \frac{\alpha}{2} - 1\right)\sin\theta_1 + \tilde{\delta}\frac{\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\qquad \text{from } \cos\theta_1 \le 1 \text{ and (40)} \\
&\le \left(1-\frac{\alpha}{2} + \left(\frac{\alpha}{2}\frac{c}{\hat{\sigma}} + \frac{\alpha}{2} - 1\right)\sin\theta_1 + \tilde{\delta}\frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2}\right)\rho_1 \\
&\qquad \text{from (23), } \sigma_1 \ge \hat{\sigma}, \text{ and } \rho_1 < \tilde{\rho}_\alpha (67) \\
&< \left(1-\frac{\alpha}{2} + \frac{\alpha}{2}\left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha\right)\sin\tilde{\phi}_\alpha + \tilde{\delta}\frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2}\right)\rho_1 \\
&\qquad \text{from } \alpha \in [1,2),\ q_\alpha < \tfrac{1}{8}, \text{ and } \theta_1 \le \psi_1 < \tilde{\phi}_\alpha \text{ (68)}.
\end{aligned}$$

By replacing $\sin\tilde{\phi}_\alpha$ with the first inequality implicit in its definition (43) and using the definition of $\tilde{\rho}_\alpha$ (47), we have

$$\rho_2 < \left(1-\frac{\alpha}{2} + \frac{\alpha}{2}q_\alpha + \frac{(1-\alpha/2-q_\alpha)\hat{\sigma}}{4}\sin\tilde{\phi}_\alpha\right)\rho_1.$$

By the first inequality implicit in (43), the definition of c (23), and $q_\alpha < \frac{1}{8}$, we have

$$\sin\tilde{\phi}_\alpha < \frac{q_\alpha}{2-(1/8)} < q_\alpha. \tag{69}$$

We can apply this bound to simplify our bound for ρ_2 as follows:

$$\begin{aligned}
\rho_2 &< \left(1 - \frac{\alpha}{2} + \frac{\alpha}{2} q_\alpha + \frac{(1-\alpha/2-q_\alpha)\hat{\sigma}}{4} q_\alpha\right)\rho_1 \\
&\le \left(1 - \frac{\alpha}{2} + \frac{\alpha}{2} q_\alpha + \frac{(1-\alpha/2)}{4} q_\alpha\right)\rho_1 && \text{using } q_\alpha > 0 \text{ and } \hat{\sigma} \le 1 \\
&\le \left(\frac{1}{2} + q_\alpha + \frac{1}{8} q_\alpha\right)\rho_1 && \text{using } \alpha \in [1,2) \\
&< \rho_1 && \text{using } q_\alpha < \tfrac{1}{8}.
\end{aligned}$$

Next, we show that $\psi_2 < \phi$. As in Subsection A.3, we define $\Delta\psi_i$ to be the angle between consecutive iterates x_i and x_{i+1}, so that $\psi_2 \le \psi_1 + \Delta\psi_1$. In addition, from (68), (46), and (18), we have $\psi_1 \le \tilde{\phi}_\alpha \le \phi \le \pi/4$. In Appendix B, we demonstrate that $\Delta\psi_1 < \pi/2$. Thus, using (68), we have

$$\sin\psi_2 \le \sin\psi_1 + \sin\Delta\psi_1 \le \frac{1-\alpha/2}{4}\sin\tilde{\phi}_\alpha + \sin\Delta\psi_1. \tag{70}$$

Since $\Delta\psi_1 \le \pi/2$, we also have

$$\sin\Delta\psi_1 \equiv \min_{\lambda\in\mathbf{R}} \|\lambda x_2 - t_1\|. \tag{71}$$

By (57) with $k = 0$, we have

$$\begin{aligned}
&\|x_2 - (1-\alpha/2)x_1\| \\
&\quad\le \left(\left\|\begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}^{-1}(t_1)\bar{C}(t_1) \\ 0 & \frac{-\alpha}{2} I\end{bmatrix} t_1\right\| + \frac{\tilde{\delta}\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\quad\le \left(\frac{\alpha}{2}\frac{c}{\sigma_1}\sin\theta_1 + \frac{\tilde{\delta}\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\qquad\qquad \text{by (16) and (23)} \\
&\quad< \frac{\alpha}{2}\left(\frac{c}{\hat{\sigma}}\sin\tilde{\phi}_\alpha + 2\frac{\tilde{\delta}\tilde{\rho}_\alpha}{\hat{\sigma}^2}\right)\rho_1 \\
&\qquad\qquad \text{by } \sigma_1 \ge \hat{\sigma},\ \sin\theta_1 < \sin\tilde{\phi}_\alpha,\ \rho_1 < \tilde{\rho}_\alpha, \text{ and } \alpha \ge 1 \\
&\quad= \frac{\alpha}{2}\left(\frac{c}{\hat{\sigma}}\sin\tilde{\phi}_\alpha + \frac{(1-\alpha/2-q_\alpha)\hat{\sigma}}{2}\sin\tilde{\phi}_\alpha\right)\rho_1 \\
&\qquad\qquad \text{by (47)} \\
&\quad\le \frac{\alpha}{2}\left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha\right)\sin\tilde{\phi}_\alpha\rho_1 \\
&\qquad\qquad \text{by } \alpha \ge 0,\ q \le 1, \text{ and } \hat{\sigma} \le 1 \\
&\quad\le \frac{\alpha}{2} q_\alpha \rho_1 \\
&\qquad\qquad \text{by the first inequality in (43).}
\end{aligned} \tag{72}$$

The inequality (72) provides a bound on $\sin \mathit{\Delta}\psi_1$ in terms of $q_\alpha(s)$:

$$\sin \mathit{\Delta}\psi_1 = \min_{\lambda \in \mathbb{R}} \|\lambda x_2 - t_1\| \leq \left\| \frac{x_2}{(1-\alpha/2)\rho_1} - t_1 \right\| < \frac{\alpha}{2(1-\alpha/2)} q_\alpha. \tag{73}$$

By substituting (73) into (70) and using (42), we find that

$$\sin \psi_2 \leq \frac{(1-\alpha/2)}{4} \sin \tilde{\phi}_\alpha + \frac{\alpha}{2(1-\alpha/2)} q_\alpha = \frac{(1-\alpha/2)}{4} \sin \tilde{\phi}_\alpha + \frac{\alpha}{8} \sin \phi. \tag{74}$$

From (69), (42), and $(1-\alpha/2) \in (0, 1/2]$, we have

$$\sin \tilde{\phi}_\alpha < \frac{8}{15} q_\alpha \leq \frac{1}{15} \sin \phi. \tag{75}$$

Therefore, we have

$$\sin \psi_2 < \left(\frac{(1-\alpha/2)}{4} \frac{1}{15} + \frac{\alpha}{8} \right) \sin \phi < \left(\frac{1}{8} \left(\frac{1}{15} \right) + \frac{\alpha}{8} \right) \sin \phi < \sin \phi. \tag{76}$$

By definition, we have $\phi \leq \frac{\pi}{4}$, from which it follows that $\sin \phi \leq \frac{1}{\sqrt{2}}$. By (76), we also have $\sin \psi_2 \leq \frac{1}{\sqrt{2}}$. This implies either $\psi_2 \leq \frac{\pi}{4}$ or $\psi_2 \geq \frac{3\pi}{4}$. However, we know that j$\psi_2 \leq \psi_1 + \mathit{\Delta}\psi_1$, and we have shown that $\psi_1 \leq \tilde{\phi}_\alpha$ and $\mathit{\Delta}\psi_1 < \frac{\pi}{2}$. By (46), we have $\tilde{\phi}_\alpha \leq \phi \leq \frac{\pi}{4}$ and therefore $\psi_2 < \frac{3\pi}{4}$. Hence, it must be the case that $\psi_2 \leq \frac{\pi}{4}$. As a result, the inequality in (76) remains valid upon removing the sine functions, that is, $\psi_2 < \phi$. This completes the proof of our claim that $x_2 \in \mathcal{I}_{s_0,\alpha}$.

To complete the anchor argument, we need to show that $\sin \theta_2 < \sin \tilde{\phi}_\alpha$. From the second row of (56) with $k = 0$, and using (144) with $x = x_1$ and (40), we have

$$\rho_2 \sin \theta_2 \leq \rho_1 (\alpha - 1) \sin \theta_1 + \alpha \delta \frac{\rho_1^2}{\sigma_1^2} < \rho_1 \left((\alpha - 1) \sin \psi_1 + \tilde{\delta} \frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2} \right),$$

where the second inequality follows from $\theta_1 \leq \psi_1$, $\alpha\delta \leq \tilde{\delta}$, $\rho_1 < \tilde{\rho}_\alpha$, and $\sigma_1 \geq \hat{\sigma}$. Using (68) and the definition of $\tilde{\rho}_\alpha$ (47), we have

$$\begin{aligned} \rho_2 \sin \theta_2 &< \rho_1 \left((\alpha - 1) \frac{(1-\alpha/2)}{4} \sin \tilde{\phi}_\alpha + \frac{(1-\alpha/2 - q_\alpha)\hat{\sigma}}{4} \sin \tilde{\phi}_\alpha \right) \\ &\leq \rho_1 \left(\alpha \frac{(1-\alpha/2)}{4} \sin \tilde{\phi}_\alpha \right) < \rho_1 \frac{(1-\alpha/2)}{2} \sin \tilde{\phi}_\alpha, \end{aligned} \tag{77}$$

with the second inequality following from $q_\alpha > 0$ and $\hat{\sigma} \leq 1$ and the third inequality a consequence of $\alpha \in [1, 2)$.

To utilize (77), we require a lower bound on ρ_2 in terms of a fraction of ρ_1. By applying the inverse triangle inequality to (72), we obtain

$$|\rho_2 - (1-\alpha/2)\rho_1| \leq \|x_2 - (1-\alpha/2)x_1\| < \frac{\alpha}{2} q_\alpha \rho_1.$$

Therefore, using (42), we obtain

$$\rho_2 \geq \rho_1 \left(\left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} q_\alpha \right) = \left(1 - \frac{\alpha}{2}\right) \rho_1 \left(1 - \frac{\alpha}{8} \sin\phi\right) > \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \rho_1,$$

where the final inequality follows from $\alpha < 2$ and $\sin\phi \leq 1$. By combining this inequality with (77), we find that

$$\rho_2 \sin\theta_2 < \rho_2 \sin\tilde{\phi}_\alpha.$$

If θ_2 is bounded above by $\frac{\pi}{2}$, this inequality is valid without the sine functions. Combining the fact that, by definition, $\theta_2 \leq \psi_2$ with the above relationship $\psi_2 < \frac{\pi}{2}$, we find that $\theta_2 < \frac{\pi}{2}$. Hence $\theta_2 < \tilde{\phi}_\alpha$ as desired.

5.4 The Induction Step

In the remainder of this proof, we provide the argument for the induction step: If (64) holds for some j, then (65) holds as well.

5.4.1 Iteration $2j+1$

We show in this subsection that if (64) holds, that is,

$$\rho_i < \tilde{\rho}_\alpha, \quad \theta_i < \tilde{\phi}_\alpha, \quad \psi_i < \phi, \qquad \text{for } i = 1, 2, \ldots, 2j, \tag{78}$$

then after the step from x_{2j} to x_{2j+1}, which is a regular Newton step, we have (65) for $\iota = 1$, that is,

$$\rho_{2j+1} < \tilde{\rho}_\alpha, \quad \theta_{2j+1} < \tilde{\phi}_\alpha, \quad \psi_{2j+1} < \phi. \tag{79}$$

Consider $k \in \{1, 2, \ldots, j\}$. In the same manner that the inequalities (151) and (153) follow from equation (149), we have the following nearly identical inequalities (80) and (81) following from equations (54) and (55):

$$\sin\theta_{2k+1} = \min_{y \in N} \|t_{2k+1} - y\| \leq \delta \frac{\rho_{2k}^2}{\sigma_{2k}^2 \rho_{2k+1}} \tag{80}$$

and

$$\left\| x_{2k+1} - \frac{1}{2} x_{2k} \right\| \leq \left(\frac{1}{2} \frac{c}{\sigma_{2k}} \sin\theta_{2k} + \delta \frac{\rho_{2k}}{\sigma_{2k}^2} \right) \rho_{2k}. \tag{81}$$

By dividing (81) by ρ_{2k} and applying the reverse triangle inequality, we have

$$\left| \frac{\rho_{2k+1}}{\rho_{2k}} - \frac{1}{2} \right| \leq \left(\frac{1}{2} \frac{c}{\sigma_{2k}} \sin\theta_{2k} + \delta \frac{\rho_{2k}}{\sigma_{2k}^2} \right). \tag{82}$$

Further, from (64), we can bound (81) as follows, for all $k = 1, 2, \ldots, j$:

(83)

$$\begin{aligned}
&\left\| x_{2k+1} - \frac{1}{2} x_{2k} \right\| \\
&< \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \tilde{\phi}_\alpha + \delta \frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2} \right) \rho_{2k} && \text{using } \sigma_{2k} \geq \hat{\sigma},\ \theta_{2k} < \tilde{\phi}_\alpha,\ \rho_{2k} < \tilde{\rho}_\alpha \\
&\leq \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \tilde{\phi}_\alpha + \frac{(1 - \alpha/2 - q_\alpha)\hat{\sigma}}{4} \sin \tilde{\phi}_\alpha \right) \rho_{2k} && \text{by (41) and (47)} \\
&< \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + \frac{(1 - \alpha/2)}{2} \right) \rho_{2k} \sin \tilde{\phi}_\alpha && \text{using } q_\alpha > 0 \text{ and } \hat{\sigma} \leq 1 \\
&< \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha \right) \rho_{2k} \sin \tilde{\phi}_\alpha && \text{using } q_\alpha < \tfrac{1}{8} \text{ and } \alpha \geq 1 \\
&\leq \frac{q_\alpha}{2} \rho_{2k} && \text{by the first part of (43).}
\end{aligned}$$

Dividing by ρ_{2k} and applying the reverse triangle inequality, we have

$$\left| \frac{\rho_{2k+1}}{\rho_{2k}} - \frac{1}{2} \right| < \frac{q_\alpha}{2}. \tag{84}$$

Therefore,

$$\frac{1 - q_\alpha}{2} \leq \frac{\rho_{2k+1}}{\rho_{2k}} \leq \frac{1 + q_\alpha}{2}. \tag{85}$$

From the right inequality, $q_\alpha < \frac{1}{8}$, and the induction hypothesis, we have

$$\rho_{2k+1} < \rho_{2k} < \tilde{\rho}_\alpha. \tag{86}$$

In particular, since k is any index in $\{1, 2, \ldots, j\}$, we have $\rho_{2j+1} < \tilde{\rho}_\alpha$. From the left inequality and (80), we have

$$\begin{aligned}
\sin \theta_{2k+1} &\leq \delta \frac{2\rho_{2k}}{\sigma_{2k}^2 (1 - q_\alpha)} \\
&< \delta \frac{2\tilde{\rho}_\alpha}{\hat{\sigma}^2 (1 - q_\alpha)} && \text{using } \rho_{2k} < \tilde{\rho}_\alpha \text{ and } \sigma_{2k} \geq \hat{\sigma} \\
&= \frac{(1 - \alpha/2 - q_\alpha)\hat{\sigma}}{2(1 - q_\alpha)} \sin \tilde{\phi}_\alpha && \text{by (41) and (47)} \\
&< \sin \tilde{\phi}_\alpha && \text{using } \hat{\sigma} \leq 1,
\end{aligned} \tag{87}$$

so that $\sin \theta_{2j+1} < \sin \tilde{\phi}_\alpha$.

In the remainder of the subsection we prove that $\psi_{2j+1} < \phi$. We consider $k \in \{2, 3, \ldots, j\}$.

We have from (78) and (22) that

$$\sigma_i \geq \hat{\sigma}, \quad i = 1, 2, \ldots, 2j, \tag{88}$$

so it follows from the definition (62) that

$$\mu_{2k-2} \geq \hat{\sigma}, \quad k = 2, 3, \ldots, j. \tag{89}$$

Since $g(x_{2k-2}) \in N$, we have

$$\begin{aligned}
\sin\theta_{2k} &= \min_{y\in N} \|t_{2k} - y\| \\
&\leq \tilde{\delta} \frac{\rho_{2k-2}^2 + \rho_{2k-1}^2}{\mu_{2k-2}^3 \rho_{2k}} && \text{from (63), with } k \leftarrow k-1 \\
&\leq 2\tilde{\delta} \frac{\rho_{2k-2}^2}{\mu_{2k-2}^3 \rho_{2k}} && \text{from (86), with } k \leftarrow k-1.
\end{aligned} \tag{90}$$

From (152) with $j = 2k-2$, we can deduce using earlier arguments that

$$\|x_{2k-2} - g(x_{2k-2})\| \leq \frac{c}{\sigma_{2k-2}} \rho_{2k-2} \sin\theta_{2k-2}.$$

By combining this bound with (63) (with $k \leftarrow k-1$), we obtain

$$\begin{aligned}
&\left\| x_{2k} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) x_{2k-2} \right\| \\
&\leq \left\| x_{2k} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) g(x_{2k-2}) \right\| + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \|x_{2k-2} - g(x_{2k-2})\| \\
&\leq \tilde{\delta} \frac{\rho_{2k-2}^2 + \rho_{2k-1}^2}{\mu_{2k-2}^3} + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \frac{c}{\sigma_{2k-2}} \rho_{2k-2} \sin\theta_{2k-2} \\
&\leq 2\tilde{\delta} \frac{\rho_{2k-2}^2}{\mu_{2k-2}^3} + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \frac{c}{\sigma_{2k-2}} \rho_{2k-2} \sin\theta_{2k-2} \\
&\qquad \text{from (86)} \\
&\leq \left[2\tilde{\delta} \frac{\tilde{\rho}_\alpha}{\hat{\sigma}^3} + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \frac{c}{\hat{\sigma}} \sin\tilde{\phi}_\alpha \right] \rho_{2k-2} \\
&\qquad \text{from (78), (89), and (88)} \\
&= \left[\frac{1}{2}\left(1 - \frac{\alpha}{2} - q_\alpha\right) + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
&\qquad \text{from (47)} \\
&\leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \left[1 - \frac{q_\alpha}{1 - \alpha/2} + \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
&\leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) \left[1 - q_\alpha + \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
&\qquad \text{from } 0 < 1 - \alpha/2 < 1 \text{ and } q_\alpha > 0 \\
&\leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) q_\alpha \rho_{2k-2} \qquad \text{from (43).}
\end{aligned} \tag{91}$$

Upon dividing by ρ_{2k-2} and applying the reverse triangle inequality, we find from the fourth line of (91) that

$$\left|\frac{\rho_{2k}}{\rho_{2k-2}} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)\right| \leq 2\tilde{\delta}\frac{\rho_{2k-2}}{\mu_{2k-2}^3} + \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)\frac{c}{\sigma_{2k-2}}\sin\theta_{2k-2}, \tag{92}$$

while from the last line of (91), we have

$$\left|\frac{\rho_{2k}}{\rho_{2k-2}} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)\right| \leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) q_\alpha. \tag{93}$$

We can restate this inequality as follows:

$$\frac{1}{2}\left(1 - \frac{\alpha}{2}\right)(1 - q_\alpha) \leq \frac{\rho_{2k}}{\rho_{2k-2}} \leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)(1 + q_\alpha), \text{ for } k = 2, 3, \ldots, j. \tag{94}$$

From the right inequality in (94), we obtain

$$\rho_{2k} \leq \left[\frac{1}{2}\left(1 - \frac{\alpha}{2}\right)(1 + q_\alpha)\right]^{k-1} \rho_2, \tag{95}$$

while by substituting the left inequality into (90) and using (89), we obtain

$$\begin{aligned} \sin\theta_{2k} &\leq 2\tilde{\delta}\frac{\rho_{2k-2}^2}{\mu_{2k-2}^3 \rho_{2k}} \\ &\leq 4\tilde{\delta}\frac{\rho_{2k-2}}{\hat{\sigma}^3}\frac{1}{(1-\alpha/2)(1-q_\alpha)} \\ &\leq 4\tilde{\delta}\frac{\rho_{2k-2}}{\hat{\sigma}^3}\frac{1}{1-\alpha/2-q_\alpha} \qquad \text{for } k = 2, 3, \ldots, j. \end{aligned} \tag{96}$$

We now define $\Delta^2\psi_i$ to be the angle between x_i and x_{i+2}. Recalling our earlier definition of $\Delta\psi_i$ as the angle between x_i and x_{i+1}, we have

$$\psi_{2j+1} \leq \psi_2 + \sum_{k=2}^{j} \Delta^2\psi_{2k-2} + \Delta\psi_{2j}. \tag{97}$$

From the fourth line of (91), we have

$$\begin{aligned} \sin\Delta^2\psi_{2k-2} &= \min_{\lambda\in\mathbb{R}} \|\lambda x_{2k} - t_{2k-2}\| \\ &= \frac{2}{(1-\alpha/2)\rho_{2k-2}} \min_{\lambda\in\mathbb{R}} \left\|\lambda x_{2k} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) x_{2k-2}\right\| \\ &\leq \frac{4\tilde{\delta}}{(1-\alpha/2)}\frac{\rho_{2k-2}}{\mu_{2k-2}^3} + \frac{c}{\sigma_{2k-2}}\sin\theta_{2k-2} \\ &\leq \frac{4\tilde{\delta}\rho_{2k-2}}{(1-\alpha/2)\hat{\sigma}^3} + \frac{c}{\hat{\sigma}}\sin\theta_{2k-2}, \end{aligned} \tag{98}$$

by (88) and (89). We show that $\Delta^2\psi_{2k-2} \le \pi/2$ for $k \in \{2,3,\ldots,j\}$ in Appendix B; this fact justifies the first equality in (98). From (95), we have

$$
\begin{aligned}
\sum_{k=2}^{j} \rho_{2k-2} &\le \sum_{k=0}^{j-2} \left[\frac{1}{2}\left(1-\frac{\alpha}{2}\right)(1+q_\alpha)\right]^k \rho_2 \\
&\le \left[1-\frac{1}{2}\left(1-\frac{\alpha}{2}\right)(1+q_\alpha)\right]^{-1} \rho_2 \\
&= \left[\frac{1}{2}-\frac{1}{2}q_\alpha+\frac{\alpha}{4}+\frac{\alpha}{4}q_\alpha\right]^{-1} \rho_2 \\
&\le \left[\frac{1}{2}-\frac{1}{2}q_\alpha+\frac{\alpha}{4}\right]^{-1} \rho_2 \\
&< \frac{2}{1+(\alpha/2)-q_\alpha}\tilde{\rho}_\alpha && \text{from (78)} \\
&= \frac{1}{2\tilde{\delta}}\frac{1-(\alpha/2)-q_\alpha}{1+(\alpha/2)-q_\alpha}\hat{\sigma}^3 \sin\tilde{\phi}_\alpha && \text{from (47).}
\end{aligned}
\tag{99}
$$

From (96) and (99) we have

$$
\begin{aligned}
\sum_{k=2}^{j} \sin\theta_{2k-2} &\le \sin\theta_2 + \sum_{k=2}^{j-1} \sin\theta_{2k} \\
&\le \sin\tilde{\phi}_\alpha + \frac{4\tilde{\delta}}{\hat{\sigma}^3}\frac{1}{1-(\alpha/2)-q_\alpha}\sum_{k=2}^{j-1}\rho_{2k-2} \\
&\le \sin\tilde{\phi}_\alpha + \frac{2}{1+(\alpha/2)-q_\alpha}\sin\tilde{\phi}_\alpha \\
&\le \sin\tilde{\phi}_\alpha + \frac{2}{11/8}\sin\tilde{\phi}_\alpha \qquad \text{since } 1+\tfrac{\alpha}{2}-q_\alpha \ge 1+\tfrac{1}{2}-\tfrac{1}{8}=\tfrac{11}{8} \\
&= \frac{27}{11}\sin\tilde{\phi}_\alpha.
\end{aligned}
\tag{100}
$$

By summing (98) over $k=2,3,\ldots,j$ and using (99) and (100), we obtain

$$
\begin{aligned}
&\sum_{k=2}^{j} \sin\Delta^2\psi_{2k-2} \\
&\quad\le \frac{4\tilde{\delta}}{(1-\alpha/2)\hat{\sigma}^3}\frac{\hat{\sigma}^3}{2\tilde{\delta}}\frac{1-(\alpha/2)-q_\alpha}{1+(\alpha/2)-q_\alpha}\sin\tilde{\phi}_\alpha + \frac{c}{\hat{\sigma}}\frac{27}{11}\sin\tilde{\phi}_\alpha \\
&\quad\le \left[\frac{2}{1+(\alpha/2)-q_\alpha}+\frac{27}{11}\frac{c}{\hat{\sigma}}\right]\sin\tilde{\phi}_\alpha \\
&\quad\le \left[\frac{16}{11}+\frac{27}{11}\frac{c}{\hat{\sigma}}\right]\sin\tilde{\phi}_\alpha,
\end{aligned}
\tag{101}
$$

where we used $q_\alpha > 0$ for the second-last inequality and $1+(\alpha/2)-q_\alpha \geq 11/8$ for the final inequality. For a bound on the term $\Delta\psi_{2j}$ of (97), we use the second-last line of (83) with $k=j$ to obtain

$$\begin{aligned}\sin\Delta\psi_{2j} &= \min_{\lambda\in\mathbf{R}}\|t_{2j}-\lambda x_{2j+1}\| \\ &= \frac{2}{\rho_{2j}}\min_{\lambda\in\mathbf{R}}\left\|\frac{1}{2}x_{2j}-\lambda x_{2j+1}\right\| \\ &\leq \frac{2}{\rho_{2j}}\left\|\frac{1}{2}x_{2j}-x_{2j+1}\right\| \\ &\leq \left(\frac{c}{\hat\sigma}+1-q_\alpha\right)\sin\tilde\phi_\alpha. \end{aligned} \tag{102}$$

The equality in (102) follows from the fact that $\Delta\psi_{2j} < \pi/2$, as shown in Appendix B.

Since each of the angles in the right-hand side of (97) is bounded above by $\pi/2$, from reasoning similar to that of Section A.3, we have

$$\sin\psi_{2j+1} \leq \sin\psi_2 + \sum_{k=2}^{j}\sin\Delta^2\psi_{2k-2} + \sin\Delta\psi_{2j}. \tag{103}$$

By substituting (76), (101), and (102) into (103), we obtain

(104)

$$\begin{aligned}\sin\psi_{2j+1} &\leq \left[\frac{1}{120}+\frac{\alpha}{8}\right]\sin\phi + \left[\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}\right]\sin\tilde\phi_\alpha + \left[\frac{c}{\hat\sigma}+1-q_\alpha\right]\sin\tilde\phi_\alpha \\ &\leq \frac{2}{7}\sin\phi + \frac{\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}}{1-q_\alpha+\frac{c}{\hat\sigma}}q_\alpha + q_\alpha \\ &\qquad \text{from (43) and } \alpha < 2 \\ &\leq \frac{2}{7}\sin\phi + \frac{\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}}{\frac{7}{8}+\frac{c}{\hat\sigma}}q_\alpha + q_\alpha \\ &\qquad \text{from } q_\alpha < \tfrac{1}{8} \\ &\leq \frac{2}{7}\sin\phi + \left(\frac{27}{11}+1\right)q_\alpha \\ &\leq \frac{2}{7}\sin\phi + \frac{38}{11}\frac{1}{8}\sin\phi \\ &< \sin\phi. \end{aligned}$$

Note that $\psi_{2j+1} \leq \psi_{2j} + \Delta\psi_{2j}$. By the induction assumption (78), we have $\psi_{2j} < \phi \leq \frac{\pi}{4}$, and from above we have $\Delta\psi_{2j} < \frac{\pi}{2}$. Hence, $\psi_{2j+1} < \frac{3\pi}{4}$. As argued after (76), this inequality combined with (104) yields $\psi_{2j+1} < \phi$, as required. Further, since $\theta_{2j+1} \leq \psi_{2j+1}$ by their definitions, inequality (87) with $k=j$ also remains valid without the sine functions, that is, $\theta_{2j+1} < \tilde\phi_\alpha$.

5.4.2 Iteration $2j+2$

We now show that

$$\rho_{2j+2} < \tilde{\rho}_\alpha, \quad \theta_{2j+2} < \tilde{\phi}_\alpha \quad \psi_{2j+2} < \phi,$$

by using some of the bounds proved above: $\rho_{2j} < \tilde{\rho}_\alpha$, $\rho_{2j+1} < \rho_{2j}$, $\theta_{2j} < \tilde{\phi}_\alpha$, $\psi_{2j} < \phi$, and $\psi_{2j+1} < \phi$. The last two assumptions guarantee that $\mu_{2j} \geq \hat{\sigma}$. The analysis in the latter part of Subsection 5.4.1 (starting from (89)) can therefore be applied for $k = j+1$. In particular, from (95) we have $\rho_{2j+2} < \tilde{\rho}_\alpha$. From (96) with $k = j+1$, using $\rho_{2j} < \tilde{\rho}_\alpha$, $\mu_{2j} \geq \hat{\sigma}$, and the definition of $\tilde{\rho}_\alpha$ (47) we have

$$\sin\theta_{2j+2} \leq 4\tilde{\delta}\frac{\rho_{2j}}{\hat{\sigma}^3(1-\alpha/2-q_\alpha)} < 4\tilde{\delta}\frac{\tilde{\rho}_\alpha}{\hat{\sigma}^3(1-\alpha/2-q_\alpha)} = \sin\tilde{\phi}_\alpha.$$

The argument for $\sin\psi_{2j+1} < \sin\phi$ is easily modified to show $\sin\psi_{2j+2} < \sin\phi$. We simply increase the upper index in the sum in (103) to $j+1$ (ignoring the final nonnegative term) to give

$$\sin\psi_{2j+2} \leq \sin\psi_2 + \sum_{k=2}^{j+1} \sin \Delta^2\psi_{2k-2}. \tag{105}$$

The bounds (95), (96), and (98) continue to hold for $k = j+1$, while (99) and (100) continue to hold if the upper bound on the summation is increased from j to $j+1$, so (101) also continues to hold if the upper bound of the summation is increased from j to $j+1$. Hence, similarly to (104), we obtain $\sin\psi_{2j+2} < \sin\phi$. Further, we can extend the argument in Appendix B that $\Delta^2\psi_{2k-2} < \pi/2$ to $k = j+1$. Adding this fact to $\psi_{2j+2} \leq \psi_{2j} + \Delta^2\psi_{2j}$ and $\psi_{2j} < \phi \leq \frac{\pi}{4}$, we have $\psi_{2j+2} < \frac{3\pi}{4}$. Repeating the argument following (76), this allows us to conclude that $\psi_{2j+2} < \phi$. Finally, since $\theta_{2j+2} \leq \psi_{2j+2}$, we similarly conclude that $\sin\theta_{2j+2} < \sin\tilde{\phi}_\alpha$ implies $\theta_{2j+2} < \tilde{\phi}_\alpha$.

Our proof of the induction step is complete.

5.5 Convergence Rate

As $j \to \infty$, we have $\rho_{2j} \to 0$ by (95) and $\rho_{2j+1} \to 0$ by (86). From (96), we also have $\theta_{2j} \to 0$. By combining these limits with (88) and (89), we see that the right-hand-side of (92) goes to zero as $j \to \infty$, and

$$\lim_{j\to\infty} \frac{\rho_{2j+2}}{\rho_{2j}} = \frac{1}{2}\left(1 - \frac{\alpha}{2}\right). \tag{106}$$

From the discussion above and (82), we also have $\lim_{j\to\infty} \frac{\rho_{2j+1}}{\rho_{2j}} = 1/2$, so that convergence is stable between the accelerated iterates.

6 Application to Nonlinear Complementarity Problems

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

$$0 \leq f(x), \;\; x \geq 0, \;\; x^T f(x) = 0.$$

Let x^* be a solution of NCP(f). We assume throughout this section and the next that f' is well defined and strongly semismooth at x^*. at x^*. We apply a nonlinear-equations reformulation to the NCP. We do not standardize the resulting equations (as we did earlier in (10) to simplify the discussions of Sections 4 and 5), as the rescaling and shifting needed to enforce this assumption would complicate this section considerably.

In this section, we tailor the convergence theorems, Theorems 1 and 2, to this formulation, 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, Definitions, and Properties

For any matrix $M \in \mathbb{R}^{p\times q}$ and any sets $\mathcal{U} \subseteq \{1, 2, \ldots, p\}$ and $\mathcal{V} \subseteq \{1, 2, \ldots, q\}$, we write $M_{\mathcal{U},\mathcal{V}}$ to denote the submatrix of M whose rows lie in $\mathcal{U}$ and columns lie in $\mathcal{V}$. The row submatrix corresponding to indices in the set $\mathcal{U}$ is denoted by $M_{\mathcal{U}}$. We denote the number of elements in any set $\mathcal{U}$ by $|\mathcal{U}|$. Let e_i denote the ith column of the identity matrix. In this section, we use the notation $\langle \cdot, \cdot \rangle$ to denote the inner product between two vectors. For any $x \in \mathbb{R}^n$, we use $\operatorname{diag} x$ 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^* as follows,

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

6.2 The Nonlinear-Equations Reformulation

We recall the nonlinear-equations reformulation Ψ (9) of the NCP (1), and consider the use of Newton's method for solving $\Psi(x) = 0$. In this section, we establish the structure of null space N and the form of 2-regularity (6) for the NCP function f with the reformulation Ψ, then tailor the local convergence results of Sections 4 and 5 to f with the reformulation Ψ.

Taking the derivative of Ψ, we have

$$\begin{aligned} \Psi_i'(x) = 2\{&(f_i(x) - \min(0, x_i + f_i(x)))e_i \\ &+ (x_i - \min(0, x_i + f_i(x)))f_i'(x)\}, \quad \text{for } \; i = 1, 2, \ldots, n. \end{aligned} \tag{107}$$

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

$$N \equiv \ker \Psi'(x^*) = \{\xi \in \mathbb{R}^n \mid f_\gamma'(x^*)\xi = 0,\ \xi_\alpha = 0\}, \tag{108}$$

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

$$N_* = \{\xi \in \mathbb{R}^n \mid \xi_\alpha = -(\operatorname{diag} f_\alpha(x^*))^{-1}(f'_{\gamma,\alpha}(x^*))^T(\operatorname{diag} x_\gamma^*)\xi_\gamma, \\ f'_{\gamma,\beta\cup\gamma}(x^*)^T(\operatorname{diag} x_\gamma^*)\xi_\gamma = 0\}. \tag{109}$$

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 (6) for Ψ at x^* and $d \in \mathbb{R}^n$ is

$$(P_{N_*}\Psi')'(x^*; d)|_N \ \text{ is nonsingular.} \tag{110}$$

By direct calculation, we have

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

where $\eta_i := \min(0, x_i + f_i(x))'(x; d)$. We can calculate this directional derivative 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

$$\frac{1}{2}(\Psi_i')'(x^*; d) = \begin{cases} \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*), & i \in \alpha, \\ (\langle f_i'(x^*), d\rangle - \min(0, d_i + \langle f_i'(x^*), d\rangle))e_i \\ \quad + (d_i - \min(0, d_i + \langle f_i'(x^*), d\rangle))f_i'(x^*), & i \in \beta, \\ \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*) + x_i^*(f_i')'(x^*; d), & i \in \gamma. \end{cases} \tag{111}$$

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 (111) as follows

(112)

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

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 $|\gamma| \times |\gamma|$. We note that the matrices Z and $Z_\perp$ are not uniquely defined by the conditions above; there are infinitely many possible choices in general for orthonormal matrices that span the subspaces in question. However the properties discussed below are independent of the particular choices for these matrices.

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

Proposition 1 *2-regularity (110) holds for $d \in \mathbb{R}^n$ at a solution x^* of $\Psi(x) = 0$ if and only if the matrix*

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\max(\langle f_i', d\rangle, -d_i)e_i - \min(\langle f_i', d\rangle, -d_i) f_i'\right]_{i\in\beta}^T \\ Z^T f_\gamma' \\ Z_\perp^T \left[(f_\gamma')'(x^*; d) + \left[\frac{1}{x_i^*}\langle f_i', d\rangle e_i^T\right]_{i\in\gamma} - \langle f'_{\gamma,\alpha}, \left(\operatorname{diag} \frac{d_j}{f_j}\right)_{j\in\alpha} f'_\alpha\rangle\right] \end{bmatrix} \tag{113}$$

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

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\max(\langle f_i', d\rangle, -d_i)e_i - \min(\langle f_i', d\rangle, -d_i) f_i'\right]_{i\in\beta}^T \\ Z^T f_\gamma' \\ Z_\perp^T (f_\gamma')'(x^*; d) \end{bmatrix} \tag{114}$$

is nonsingular.

Proof Consider any $d \in \mathbb{R}^n$. The claim that $(P_{N_*}\Psi')'(x^*; d)|_N$ is nonsingular for some $d \in \mathbb{R}^n$ (110) 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 (108) and (112) that

(115)
$$\frac{1}{2}(\Psi_i')'(x^*;d)v = \begin{cases} d_i\langle f_i', v\rangle, & i \in \alpha \\ \max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle, & i \in \beta \\ \langle f_i', d\rangle v_i + x_i^*\langle (f_i')'(x^*;d), v\rangle, & i \in \gamma. \end{cases}$$

Since N_* is defined in (109) to have the form $\{\xi \in \mathbb{R}^n \,|\, A\xi = 0\}$ for some matrix A, we have that $P_{N_*}w = 0$ if and only if $w = A^T z$ for some z. That is,

$$\frac{1}{2}(\Psi')'(x^*;d)v = \begin{bmatrix} \operatorname{diag} f_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ (\operatorname{diag} x_\gamma^*)f'_{\gamma,\alpha} & (\operatorname{diag} x_\gamma^*)f'_{\gamma,\beta} & (\operatorname{diag} x_\gamma^*)f'_{\gamma,\gamma} \end{bmatrix} \begin{bmatrix} z_\alpha \\ z_\beta \\ z_\gamma \end{bmatrix}, \tag{116}$$

for some $z \in \mathbb{R}^n$. Hence, by matching components from this expression and (115), we have that $P_{N_*}(\Psi')'(x^*;d)v = 0$ if for some $z \in \mathbb{R}^n$ we have

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

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

$$0 = \max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle, \quad i \in \beta, \tag{117a}$$

$$\begin{aligned} & \langle f_i', d\rangle v_i + x_i^*\langle (f_i')'(x^*;d), v\rangle - x_i^*\left\langle f'_{i,\alpha}, \operatorname{diag}\,(d_j/f_j)_{j\in\alpha}\,\langle f'_\alpha, v\rangle\right\rangle \\ & \qquad = x_i^*\left[f'_{i,\beta}z_\beta + f'_{i,\gamma}z_\gamma\right], \quad i \in \gamma. \end{aligned} \tag{117b}$$

Using the definition of the orthonormal matrix Z, we can rewrite (117b) as follows:

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

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

$$Z_\perp^T\left[\frac{1}{x_i^*}\langle f_i', d\rangle e_i^T + (f_i')'(x^*;d) - \langle f'_{i,\alpha}, \operatorname{diag}\left(\frac{d_j}{f_j}\right)_{j\in\alpha} f'_\alpha\rangle\right]_{i\in\gamma} v = 0. \tag{118}$$

Since $v \in N$, we have from (108) that

$$v_\alpha = 0, \tag{119a}$$

$$f'_{\gamma,\alpha}v_\alpha + f'_{\gamma,\beta}v_\beta + f'_{\gamma,\gamma}v_\gamma = 0. \tag{119b}$$

The second condition (119b) is equivalent to

$$\begin{bmatrix} Z^T \\ Z_\perp^T \end{bmatrix}\left[f'_{\gamma,\alpha}\; f'_{\gamma,\beta}\; f'_{\gamma,\gamma}\right]v = 0. \tag{120}$$

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 (120) does not add any information and can be dropped. Hence, we can write (119) equivalently as

$$v_\alpha = 0, \qquad Z^T \left[f'_{\gamma,\alpha}\ f'_{\gamma,\beta}\ f'_{\gamma,\gamma}\right] v = 0. \tag{121}$$

By gathering the conditions equivalent to $v \in N$ and $P_{N_*}(\Psi')'(x^*;d)v = 0$, namely, (117a), (118), and (121), we have

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

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 (113) is nonsingular. For $d \in N$, we have by the definition of N (108) 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 (114).

Recall that Ψ (9) is 2^1-regular (6) at x^* if $(P_{N_*}\Psi')'(x^*;d)|_N$ is nonsingular for some d in N, that is, if the matrix (114) 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 (114) 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 (38) and (39) 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 (113) and (114) simplify to more familiar regularity conditions in special cases of the NCP.

Nondegenerate NCP. Considering now the special 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-regularity for $d \in N$ at the solution x^* if and only if*

$$P_{N^f_{*\gamma}} (f'_{\gamma,\gamma})'(x^*; d)|_{N^f_\gamma} \tag{122}$$

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 (114), 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) and the columns of $\bar{Z}$ span range $(f'_{\gamma,\gamma})^T$. We have that $\bar{Z} \in \mathbb{R}^{|\gamma| \times r}$ and that $\bar{Z}_\perp \in \mathbb{R}^{|\gamma| \times (|\gamma| - r)}$ and, by the fundamental theorem of linear algebra, that $\left[\bar{Z} \,|\, \bar{Z}_\perp\right]$ is orthogonal. Specializing 2-regularity for $d \in N$ (114) to the case of $\beta = \emptyset$, we have that 2-regularity is equivalent to nonsingularity of the following matrix for some $d \in N$:

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

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 $Z_\perp$ defined earlier 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 (122).

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 (108) that $N = \{0\}$, so that x^* is a nonsingular solution and Theorem 3 does not apply.

Consider now the case in which $\alpha = \beta = \emptyset$ but $f'(x^*)$ has rank less than n—essentially the case of degenerate 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 (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 $\mathbb{R}^n$*

$$\{e_i\}_{i\in\alpha} \cup \{f_i'(x^*)\}_{i\in\gamma} \cup \{\langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*)\}_{i\in\beta_1} \cup \{\langle f_i'(x^*), d\rangle f_i'(x^*) + d_i e_i\}_{i\in\beta_2}, \tag{123}$$

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

$$\beta_1(d) := \{i \in \beta \,|\, \langle f_i'(x^*), d\rangle > -d_i\}, \tag{124a}$$

$$\beta_2(d) := \{i \in \beta \,|\, \langle f_i'(x^*), d\rangle \le -d_i\}, \tag{124b}$$

is linearly independent, then 2-regularity (113) 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 (123) 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 (113) reduces to

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

By partitioning the index set β according to (124), we see that nonsingularity of this matrix is equivalent to linear independence of the vectors (123). This proves the converse implication, since it assumes that $f'_{\gamma,\beta\cup\gamma}(x^*)$. The first implication follows by noting that linear independence of the set (123) implies that $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank.

As discussed at the end of Section 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 (123) for the partition (β_1, β_2) of β defined in (124) 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]. It requires linear independence of the vectors (123) for every partition (β_1, β_2) of β for some $d \in \mathbb{R}^n$.

6.4 "Directional Denseness" of the Starlike Domain.

In this subsection, we give sufficient conditions for the starlike domain of convergence $\mathcal{R}$ (33) (or $\mathcal{R}_\alpha$ (50)), to be "directionally dense" at the solution x^*.

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

$$\text{there exists a positive number } C_t \text{ such that} \quad \text{if } \rho < C_t \text{ then } x = x^* + \rho t \in \mathcal{R}. \tag{125}$$

A direction t satisfies (125) if and only if t is not an *excluded direction*, as defined in Section 2.

We recall the characterization of the excluded directions of $\mathcal{R}$ from (35): A direction $t \in \mathcal{S}$ is excluded if and only if one of the following conditions is true:

$$t \in \Pi_0^{-1}(0), \tag{126a}$$

$$g(t) = 0, \text{ or} \tag{126b}$$

$$g(t)/\|g(t)\| \in \Pi_0^{-1}(0). \tag{126c}$$

In the following, we tailor the definitions of Π_0 and $g(t)$ to our application. (We do not use the standardizing assumptions (10) in the following definitions.) Consider the first and third conditions (126a) and (126c). We recall the definition of $\Pi_0 : \mathbb{R}^n \to \mathbb{R}$,

$$\Pi_0(d) := \det(P_{N_*}\Psi')'(x^*; d)|_N, \qquad \text{for } d \in \mathbb{R}^n. \tag{127}$$

Observe that the condition $d \notin \Pi_0^{-1}(0)$ is equivalent to the 2-regularity condition (110) for $d \in \mathbb{R}^n$, which is itself equivalent to nonsingularity of the matrix (113) by Proposition 1. Further, if $d \in N$, the condition $d \notin \Pi_0^{-1}(0)$ is equivalent to nonsingularity of the simpler matrix (114).

Now consider the second condition (126b). For $x \in \mathbb{R}^n$ with $\Pi_0(x-x^*) \neq 0$ and $\|x - x^*\|$ sufficiently small, recall 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

$$\begin{aligned} g(x - x^*) = \rho g(t) &= P_N(x - x^*) \\ &+ ((P_{N_*}\Psi')'(x^*; t)|_N)^{-1}(P_{N_*}\Psi')'(x^*; t)|_{N_\perp} P_{N_\perp}(x - x^*), \end{aligned} \tag{128}$$

for $x = x^* + \rho t$, $\rho = \|x - x^*\|$, and $t \in \mathcal{S}$. As in (36), we have

$$g(d) = 0 \Leftrightarrow (P_{N_*}\Psi')'(x^*; d)d = 0, \qquad \text{for } d \in \mathbb{R}^n \setminus \Pi_0^{-1}(0). \tag{129}$$

From (112) and dividing the set β into $\beta_1(d)$ and $\beta_2(d)$ (124) for $d \in \mathbb{R}^n$, we have

$$\frac{1}{2}(\Psi_i')'(x^*; d)d = \begin{cases} 2d_i\langle f_i'(x^*), d\rangle, & i \in \alpha, \\ 2d_i\langle f_i'(x^*), d\rangle, & i \in \beta_1(d), \\ -d_i^2 - \langle f_i'(x^*), d\rangle^2, & i \in \beta_2(d), \\ 2d_i\langle f_i'(x^*), d\rangle + x_i^*\langle (f_i')'(x^*; d), d\rangle, & i \in \gamma. \end{cases} \tag{130}$$

In order to express $(P_{N_*}\Psi')'(x^*;d)d = 0$ in terms of f, 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 (116). 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

$$2d_i\langle f_i'(x^*), d\rangle = f_i z_i, \qquad i \in \alpha, \tag{131a}$$

$$2d_i\langle f_i'(x^*), d\rangle = 0, \qquad i \in \beta_1(d), \tag{131b}$$

$$d_i^2 + \langle f_i'(x^*), d\rangle^2 = 0, \qquad i \in \beta_2(d), \tag{131c}$$

$$2d_i\langle f_i'(x^*), d\rangle + x_i^*\langle (f_i')'(x^*;d), d\rangle = x_i^*\langle f_i', z\rangle, \qquad i \in \gamma. \tag{131d}$$

If $\beta \neq \emptyset$ and $f'_\beta \not\equiv 0$, then (131b) and (131c) fail almost surely. This is because, for any $d \in \mathbb{R}^n$, d_i is almost surely nonzero for $i = 1, 2, \cdots, n$, and, if $f'_\beta \not\equiv 0$, then $\langle f_i', d\rangle$ is almost surely nonzero for $i \in \beta$. In this case, we have $(P_{N_*}\Psi')'(x^*;d)d \neq 0$ almost surely for $d \in \mathbb{R}^n$. If $\beta = \emptyset$, the conditions (131) can be simplified as follows. Solve (131a) for z_α. Substituting z_α into (131d), we find that $(P_{N_*}\Psi')'(x^*;d)d = 0$ for $d \in \mathbb{R}^n$ if and only if there is some $z_\gamma \in \mathbb{R}^{|\gamma|}$ that solves

$$2\left(\operatorname{diag}\frac{d_\gamma}{x_\gamma^*}\right)\langle f_\gamma'(x^*), d\rangle + \langle (f_\gamma')'(x^*;d), d\rangle - f'_{\gamma,\alpha}(x^*)z_\alpha = f'_{\gamma,\gamma}(x^*)z_\gamma. \tag{132}$$

Since, by assumption, the (left) null space N_* is nontrivial, $\ker(f'_{\gamma,\gamma}(x^*))^T$ (109) is nontrivial. Hence, the complementary space $\operatorname{range}(f'_{\gamma,\gamma}(x^*))$ must be a strict subspace of $\mathbb{R}^{|\gamma|}$. Thus, equation (132) is solvable only if the left-hand side, which is an element of $\mathbb{R}^{|\gamma|}$, lies in the subspace spanned by $\operatorname{range}(f'_{\gamma,\gamma}(x^*))$ as is required by the right-hand side. Although counterexamples can be constructed, it seems likely that this containment will typically fail for almost all directions $d \in \mathbb{R}^n$. Under this assumption, $(P_{N_*}\Psi')'(x^*;d)d \neq 0$ almost surely for $d \in \mathbb{R}^n$. (By positive homogeneity, this is equivalent to $(P_{N_*}\Psi')'(x^*;t)t \neq 0$ almost surely for $t \in \mathcal{S}$.)

In summary, the starlike domain of convergence $\mathcal{R}$ is *directionally dense* (8) at the solution x^* if *(1)* nonsingularity of (113) holds for almost every $d = t \in \mathcal{S}$, *(2)* for almost every $d \in \mathbb{R}^n$, the system of equations (131) fails to have a solution $z \in \mathbb{R}^n$, and *(3)* nonsingularity of (114) 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. Condition *(3)* arises from (126c), which requires almost surely that $g(t)/\|g(t)\| \notin \Pi_0^{-1}(0)$ for $t \in \mathcal{S}$. If we assume that $N \cap \Pi_0^{-1}(0) = \{0\}$, then condition *(3)* is trivially satisfied by the fact that $\operatorname{range} g = N$. The assumption $N \cap \Pi_0^{-1}(0) = \{0\}$ appears in Section 1 under the name $2^\forall$-regularity (4). As discussed in Section 1, $2^\forall$-regularity is a strong form of 2-regularity. In particular, it implies isolation of the solution. However, by assuming $2^\forall$-regularity, we can write the conditions ensuring directional denseness of the starlike domain of convergence entirely in terms of f:

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 (38) and (39)) applied to the nonlinear-equations reformulation $\Psi(x)$ of NCP(f) is directionally dense if the following conditions hold:

(i) the matrix (113) is nonsingular for almost every $d \in \mathbb{R}^n$,
(ii) the system of equations (131) has no solution $z \in \mathbb{R}^n$ for almost every $d \in \mathbb{R}^n$, and
(iii) the matrix (114) is nonsingular for every $d \in N \setminus \{0\}$.

6.5 A partial equivalence between b-regularity and 2^{ae}-regularity.

We recall the definition of 2^{ae}-regularity from Section 2.

Definition 9 ($\mathbf{2^{ae}}$-regularity) 2^{ae}*-regularity* holds for Ψ at x^* if $(P_{N_*}\Psi')'(x^*;d)|_N$ is nonsingular for almost every $d \in N$.

For the NCP equation reformulation Ψ, 2^{ae}-regularity is equivalent to nonsingularity of (114) for almost every $d \in N$. The definition of b-regularity is as follows.

Definition 10 (b-regularity) The solution x^* satisfies *b-regularity* if for every partition (β_1, β_2) of β, the vector set

$$\{f_i'(x^*)\}_{i\in\beta_1\cup\gamma} \cup \{e_i\}_{i\in\beta_2\cup\alpha}$$

is linearly independent.

b-regularity is the one of the weakest known conditions implying superlinear convergence of a nonsmooth Newton method [12, p. 398]. It is also one of the weakest known conditions providing an error bound for the NCP problem [11, p. 415]. Further, it is well-known that b-regularity implies isolation of the solution. (See, for example, Corollary 3.3.9 of [6].) In Proposition 1 of [11], Izmailov and Solodov prove that b-regularity implies 2^T-regularity (7), a condition that is itself sufficient for isolation of the solution.

When $|\beta| \neq 1$, there is no equivalence between 2^{ae}-regularity and b-regularity and neither condition implies the other. (As shown in Table 1, the solutions of several test problems satisfy 2^{ae}-regularity but fail b-regularity. Conversely, the problem doubleknot satisfies b-regularity but not 2^{ae}-regularity. In fact, doubleknot satisfies strong regularity.)

The following propositions demonstrate the equivalence of 2^{ae}-regularity and b-regularity when $|\beta| = 1$ and $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank. Note that when the latter condition holds, $Z_\perp$ is null and we can take $Z = I$. The 2-regularity matrices (113) and (114) thus become identical to the following matrix:

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[\langle f_i'(x^*), d\rangle e_i^T + d_i f_i'(x^*)^T\right]_{i\in\beta_1} \\ -\left[\langle f_i'(x^*), d\rangle f_i'(x^*)^T + d_i e_i^T\right]_{i\in\beta_2} \\ f_\gamma' \end{bmatrix}, \tag{133}$$

for $\beta_1 = \beta_1(d)$ and $\beta_2 = \beta_2(d)$ (see (124)).

Proposition 2 *Suppose that* x^* *is a b-regular solution of the NCP and that* $|\beta| = 1$*. Then* $f'_{\gamma,\gamma}(x^*)$ *is nonsingular and 2-regularity (which is equivalent to full rank of (133) for* $\beta_1 = \beta_1(d)$ *and* $\beta_2 = \beta_2(d)$*) holds for almost every* $d \in \mathbb{R}^n$ *and, in addition, for all* $d \in N \setminus \{0\}$*.*

Proof Let i_0 be the index in $\{1, 2, \ldots, n\}$ such that $\beta = \{i_0\}$. By taking $\beta_1 = \emptyset$ and $\beta_2 = \beta = \{i_0\}$ in Definition 10, we have immediately that $f'_{\gamma,\gamma}(x^*)$ has full rank.

We note first from (124) that either $f'_{i_0}(x^*) = -e_{i_0}$, or else the plane $\{d \,|\, \langle f'_{i_0}(x^*) + e_{i_0}, d\rangle = 0\}$ partitions $\mathbb{R}^n$ into two half-spaces. In the first case, we have $\beta_2(d) = \{i_0\}$ for all $d \in \mathbb{R}^n$, while in the second case we have that $\{d \,|\, \beta_1(d) = \{i_0\}\}$ and $\{d \,|\, \beta_2(d) = \{i_0\}\}$ are both half-spaces of $\mathbb{R}^n$. In any case, to prove that (133) is nonsingular for almost all d, it suffices to prove that the following two matrices are nonsingular for almost all $d \in \mathbb{R}^n$:

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \langle f'_{i_0}(x^*), d\rangle e_{i_0}^T + d_{i_0} f'_{i_0}(x^*)^T \\ f'_\gamma \end{bmatrix}, \quad \begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \langle f'_{i_0}(x^*), d\rangle f'_{i_0}(x^*)^T + d_{i_0} e_{i_0}^T \\ f'_\gamma \end{bmatrix}. \tag{134}$$

We deal first with the special case of $f'_{i_0}(x^*) = \tau e_{i_0}$ for some τ. Clearly $\tau \neq 0$ by b-regularity. By substituting into (134), we obtain

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ 2\tau d_{i_0} e_{i_0}^T \\ f'_\gamma \end{bmatrix}, \quad \begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ (\tau^2 + 1) d_{i_0} e_{i_0}^T \\ f'_\gamma \end{bmatrix}.$$

Both of these matrices are obviously nonsingular whenever $d_{i_0} \neq 0$, that is, for almost all $d \in \mathbb{R}^n$.

When $f'_{i_0}(x^*)$ is not parallel to e_{i_0}, there exists a $d \in \mathbb{R}^n$ such that $d_{i_0} = 1$ and $\langle f'_{i_0}(x^*), d\rangle = 0$. For this d we have by b-regularity that both matrices (134) are nonsingular. Noting that the determinants of the matrices (134) are polynomials in the elements of d, we have that both matrices are nonsingular for almost all $d \in \mathbb{R}^n$, as desired.

For the final claim, we show that the two matrices in (134) are nonsingular for all $d \in N \setminus \{0\}$. (This part of the proof closely follows that of [12, Proposition 4.3].) By b-regularity, and defining the space L as the range of the set

$$\{e_i\}_{i\in\alpha} \cup \{f'_i(x^*)\}_{i\in\gamma}, \tag{135}$$

we have that $e_{i_0} \notin L$ and $f'_{i_0}(x^*) \notin L$. Noting that $L \equiv N_\perp$ from (135) and (108), we have for all $d \in N \setminus \{0\}$ that

$$\langle e_{i_0}, d\rangle = d_{i_0} \neq 0, \quad \langle f'_{i_0}(x^*), d\rangle \neq 0. \tag{136}$$

If one of the matrices in (134) is singular for some $d \in N \setminus \{0\}$, we have that either

$$\langle f'_{i_0}(x^*), d\rangle e_{i_0} + d_{i_0} f'_{i_0}(x^*) \in L \ \text{ or } \ \langle f'_{i_0}(x^*), d\rangle f'_{i_0}(x^*) + d_{i_0} e_{i_0} \in L,$$

so by taking inner products with $d \in N$ and using $L \equiv N_\perp$, we obtain either

$$2\langle f'_{i_0}(x^*), d\rangle d_{i_0} = 0 \ \text{ or } \ \langle f'_{i_0}(x^*), d\rangle^2 + d_{i_0}^2 = 0.$$

In either case, at least one of d_{i_0} and $\langle f'_{i_0}(x^*), d\rangle$ is zero, contradicting (136) and proving the claim.

We prove a partial converse of this result by modifying and extending the argument given by Izmailov and Solodov [12, p. 400].

Proposition 3 *Suppose that x^* is a solution of the NCP with $|\beta| = 1$. Suppose also that $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank $|\gamma|$ and 2^{ae}-regularity holds at x^*. Then x^* is b-regular.*

Proof We show the contrapositive: *Consider a solution x^* of the NCP with $|\beta| = 1$ and assume that $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank. Suppose that x^* is not b-regular. Then it cannot be true that (133) is nonsingular for almost every $d \in N$.*

As above, let L be the range of the set (135). Let $\beta = \{i_0\}$, and assume that b-regularity does not hold. If (135) is rank deficient then (133) is singular and we are done. Otherwise, since b-regularity fails we must have either $f'_{i_0} \in L$ or $e_{i_0} \in L$ (or possibly both).

Suppose first that $f'_{i_0} \in L$. Since $L \equiv N_\perp$ by (108) we have for all $d \in N$ that $\langle f'_{i_0}(x^*), d\rangle = 0$. Therefore from (124), we have

$$d_{i_0} > 0 \ \Leftrightarrow\ \beta_1(d) = \{i_0\}, \quad \beta_2(d) = \emptyset \tag{137a}$$

$$d_{i_0} \le 0 \ \Leftrightarrow\ \beta_1(d) = \emptyset, \quad \beta_2(d) = \{i_0\}. \tag{137b}$$

In the case (137a), (133) is singular since

$$\langle f'_{i_0}(x^*), d\rangle e_{i_0} + d_{i_0} f'_{i_0}(x^*) = d_{i_0} f'_{i_0}(x^*) \in L.$$

Hence for (133) to be nonsingular for almost all $d \in N$, we must have (137b) satisfied for almost all $d \in N$. Since N is a subspace, this fact implies that $d_{i_0} = 0$ for all $d \in N$. We therefore have that

$$\langle f'_{i_0}(x^*), d\rangle f'_{i_0}(x^*) + d_{i_0} e_{i_0} = 0 \tag{138}$$

for almost all $d \in N$, so that (133) is singular for almost all $d \in N$.

We now consider the case of $e_{i_0} \in L$. Since $L \equiv N_\perp$, we have that $d_{i_0} = 0$ for almost all $d \in N$. Thus from (124) we have

$$\langle f'_{i_0}(x^*), d\rangle > 0 \ \Leftrightarrow\ \beta_1(d) = \{i_0\}, \quad \beta_2(d) = \emptyset \tag{139a}$$

$$\langle f'_{i_0}(x^*), d\rangle \le 0 \ \Leftrightarrow\ \beta_1(d) = \emptyset, \quad \beta_2(d) = \{i_0\}. \tag{139b}$$

In case (139a), (133) is singular since

$$\langle f'_{i_0}(x^*), d\rangle e_{i_0} + d_{i_0} f'_{i_0}(x^*) = \langle f'_{i_0}(x^*), d\rangle e_{i_0} \in L.$$

Hence for (123) to be linearly independent, we must be in the case (139b) for almost all $d \in N$. Since N is a subspace, we thus have $\langle f'_{i_0}(x^*), d\rangle = 0$ for almost every $d \in N$. This fact implies that (138) holds for almost all $d \in N$, so that (133) is singular for almost all $d \in N$.

We have shown that when $|\beta| = 1$ and $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank then 2^{ae}-regularity is equivalent to b-regularity. However, when $f'_{\gamma,\beta\cup\gamma}(x^*)$ is rank deficient, 2^{ae}-regularity may hold, but b-regularity necessarily fails.

Finally, we mention that b-regularity of a singular solution implies the regularity requirement $(P_{N_*}\Psi')'(x^*; d)d \neq 0$ (129) for d restricted to $N \setminus \{0\}$. This observation can be justified as follows. By b-regularity and $N \neq \{0\}$, we necessarily have $\beta \neq \emptyset$. Using the definition of N (108), we have $N_\perp = \text{range}\,\{(e_i)_{i\in\alpha}, (f'_\gamma)^T\}$. From b-regularity, for any $i \in \beta$, we have $e_i \notin N_\perp$ and $f'_i \notin N_\perp$. Hence, for $d \in N \setminus \{0\}$, we have $d_i \neq 0$ and $\langle f'_i, d\rangle \neq 0$, ensuring that conditions (131b) and (131c) fail, regardless of the partition (β_1, β_2) of β. Thus the entire condition (131), which is equivalent to $(P_{N_*}\Psi')'(x^*; d)d = 0$, fails for $d \in N \setminus \{0\}$.

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 C. Properties of the problems are shown in Table 1. If the problem has more than one default starting point/solution pair, the pair's number is given following the problem name. These starting points and solutions are listed in Table 3. The convergence rate shown 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 (7), 2^{ae}-regularity is defined in (5). For a definition of b-regularity, see Definition (3.3.10) of [6].)

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. This is consistent with the fact that 2^T-regularity is sufficient for isolation. In contrast, 2^{ae}-regularity, that is, 2-regularity for almost every $d \in N$, is not sufficient for isolation of the solution. In fact, it holds for quadknot at this solution. For quadknot, we observe convergence to this solution from arbitrary nearby starting points. For affknot1, 2^{ae}-regularity fails, 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 (108) gives a direction for which 2-regularity holds. Specifically, for affknot1, $N = \{ke_2 \,|\, k \in \mathbb{R}\}$. 2-regularity along $d = ke_2$ fails if $k \geq 0$ and holds if $k < 0$. From starting points $x_0 = (x^1, x^2)$, if $x^2 < 1$, we observe convergence to the solution $x^* = (0, 1)$ with rate $\frac{1}{2}$, while if $x^2 > 1$, we observe one-step convergence to the solution $(0, x^2)$.

Despite failing 2-regularity for almost every $d \in N$, the problem affknot1 is 2-regular for almost every $d \in \mathbb{R}^n$. Only affknot1, quad1, and quad2 satisfy 2-regularity for directions in N (or $\mathbb{R}^n$) having positive measure less than 1. As we have discussed, 2-regularity holds for affknot1 on half of the directions in 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 solution from nearby points occurs with two different rates. The first starting points for quad1 and quad2 demonstrate convergence along a direction satisfying 2-

Table 1 Convergence rate of Newton's Method on Ψ for the Simple NCP test problems, showing regularity properties. ($\bullet$ = property satisfied, blank = property not satisfied, — = property not applicable.)

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

regularity, while the second starting points demonstrate convergence along a direction failing 2-regularity. As expected, the convergence rate is $\frac{1}{2}$ for the first starting points, and slower for the second starting points.

Of these problems, all but quarquad2, quarp,2, and quarn satisfy 2-regularity (110) for some $d \in \mathbb{R}^n$ and most satisfy 2-regularity for almost every $d \in N$ as can be seen from Table 1. Further, most of these 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 2, we report the numbers of iterations required for local convergence of Newton's method and the Accelerated-Newton method of Section 5 for the subset of 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 $\frac{1}{2}$ for Newton's method but do not satisfy 2^{ae}-regularity. Despite the absence of 2^{ae}-regularity, the acceleration technique of Section 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 $\|\Psi(x)\| \le 10^{-11}$.

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. 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.

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

The final column of Table 2 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.

A Convergence of Newton's Method: Details of Proof

We present here the remaining details of the proof of Theorem 1. The analysis follows that of Griewank [8] closely, but various aspects of it are referred to in our discussion of the accelerated Newton's method in Section 5, so we state it in full here.

We pick up the thread from the end of Section 4.

A.1 The Form of a Newton Step from $x \in \bar{\mathcal{R}}$

The content of this subsection is taken directly from [8] (with k set to 1); we include it here for completeness and readability of this section and for further reference in Section 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 (19) to the point $\bar{x}$, where

$$\bar{x} := x - F'(x)^{-1}F(x). \tag{140}$$

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 dependence of various matrix quantities on x. Using positivity of σ and (13), it can be checked that the following expressions from [8] are also true here.

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

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

$$G^{-1}(x) = \rho^{-1}\bar{B}^{-1}(t) + \sigma^{-2}O(\rho^0) = \sigma^{-2}O(\rho^{-1}), \tag{141}$$

(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 (13) to aggregate the order terms, as in Griewank [8], we have

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

Due to (13), (141), and the positivity of σ, we have

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

Hence,

$$\begin{aligned} &F'(x)^{-1}F(x) = \\ &\begin{bmatrix} \frac{1}{2}I + \|G^{-1}\|O(\rho^2) & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) + \sigma^{-2}O(\rho) + \|G^{-1}\|O(\rho^2) \\ O(\rho) + \|G^{-1}\|O(\rho^3), & I + O(\rho) + \|G^{-1}\|\rho^2 \end{bmatrix} x. \end{aligned} \tag{142}$$

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

$$F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & I \end{bmatrix} x - e(x), \tag{143}$$

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

$$\|e(x)\| \le \delta \frac{\|x\|^2}{\sigma(x/\|x\|)^2} = \delta \frac{\rho^2}{\sigma^2}, \tag{144}$$

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 (13) which, as we have already noted, are finite. The definition of $r(t)$ (34) uses this value of δ.

Using (143), we have

$$\bar{x} = x - F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & \frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x + e(x) = \frac{1}{2}g(x) + e(x), \tag{145}$$

where $g(x)$ is defined in (32). 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.

A.2 Entering $\mathcal{W}_{s_0}$

Denoting the sequence of Newton iterates by $\{x_j\}_{j\geq 0}$, we use the following associated notation in the remainder of this section:

$$\rho_j = \|x_j\|, \quad t_j = x_j/\rho_j, \quad \sigma_j = \sigma(t_j), \quad s_j = g(x_j)/\|g(x_j)\|. \tag{146}$$

For $s \in \mathcal{S}$, let $\psi_j(s)$ denote the angle between $t_j = x_j/\rho_j$ and s, that is,

$$\psi_j(s) = \cos^{-1} t_j^T s. \tag{147}$$

We show that if $x_0 = \rho_0 t_0 \in \mathcal{R}$, then

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

so that $x_1 \in \mathcal{W}_{s_0}$. In the next subsection, we show that all subsequent iterates remain in $\bar{\mathcal{R}}$ and converge linearly to x^*.

In the analysis below, we make repeated use of the following relations. Let $\omega_{v,s}$ denote the angle between two vectors $v \in \mathbb{R}^n$ and $s \in \mathcal{S}$. This angle must lie in the range $[0,\pi]$. If $\omega_{v,s} \leq \pi/2$, we have

$$\sin\omega_{v,s} = \min_{\lambda\in\mathbb{R}} \|\lambda v - s\|, \tag{148}$$

as well as $\sin\omega_{v,s} \geq 0$. If v is a linear subspace rather than a vector, $\omega_{v,s} \leq \pi/2$ trivially, and the above relations hold.

By applying (145) to the Newton step from x_j to x_{j+1}, we have

$$\left\|x_{j+1} - \frac{1}{2}g(x_j)\right\| = \left\|x_j - F'(x_j)^{-1}F(x_j) - \frac{1}{2}g(x_j)\right\| = \|e(x_j)\| \leq \delta\frac{\rho_j^2}{\sigma_j^2}, \tag{149}$$

By setting $j = 0$ in (149) and using $x_0 \in \mathcal{R}$, we have

$$\sin\psi_1(s_0) = \min_{\lambda\in\mathbb{R}} \left\|\lambda x_1 - \frac{g(x_0)}{\|g(x_0)\|}\right\| \leq \left(\frac{1}{2}\|g(t_0)\|\right)^{-1} \delta\frac{\rho_0}{{\sigma_0}^2}. \tag{150}$$

The equality in (150) is a consequence of (148), provided that $\psi_1(s_0) \leq \pi/2$. (We verify the latter fact in Appendix B.) By the third part of the definition of $r(t)$ (34), we have from (150) that $\sin\psi_1(s_0) < \sin\hat{\phi}(s_0)$ and therefore $\psi_1(s_0) < \hat{\phi}(s_0)$. It remains to show that $\rho_1 < \hat{\rho}(s_0)$.

Let θ_{j+1} denote the angle between the iterate x_{j+1} and the null space N. By dividing (149) by ρ_{j+1}, we obtain

$$\sin\theta_{j+1} = \min_{y\in N} \|t_{j+1} - y\| \leq \delta\frac{\rho_j^2}{\sigma_j^2\rho_{j+1}}, \text{ for } j = 0,1,2,\ldots. \tag{151}$$

The equality in (151) is valid because N is a linear subspace. (For the starting point we have only the trivial upper bound $\sin\theta_0 \leq 1$.) By the definition of g (32), we have

$$\begin{aligned} x_j - g(x_j) &= \begin{bmatrix} x_j\cos\theta_j \\ x_j\sin\theta_j \end{bmatrix} - \begin{bmatrix} I & \bar{B}^{-1}(t_j)\bar{C}(t_j) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_j\cos\theta_j \\ x_j\sin\theta_j \end{bmatrix} \\ &= \begin{bmatrix} -\bar{B}^{-1}(t_j)\bar{C}(t_j)x_j\sin\theta_j \\ x_j\sin\theta_j \end{bmatrix}. \end{aligned} \tag{152}$$

By combining (149) and (152), we obtain

$$\begin{aligned} \left\|x_{j+1}-\frac{1}{2}x_j\right\| &\le \left\|\frac{1}{2}(x_j-g(x_j))\right\|+\left\|x_{j+1}-\frac{1}{2}g(x_j)\right\| \\ &\le \frac{1}{2}\left(\|-\bar{B}^{-1}(t_j)\bar{C}(t_j)\|+1\right)\|x_j\sin\theta_j\|+\delta\frac{{\rho_j}^2}{\sigma_j^2} \\ &\le \left(\frac{1}{2}\frac{\|\bar{C}(t_j)\|+\sigma_j}{\sigma_j}\sin\theta_j+\delta\frac{\rho_j}{\sigma_j^2}\right)\rho_j \\ &\le \left(\frac{1}{2}\frac{c}{\sigma_j}\sin\theta_j+\delta\frac{\rho_j}{\sigma_j^2}\right)\rho_j, \end{aligned} \tag{153}$$

where the final inequality follows from (23). Dividing by ρ_j and applying the inverse triangle inequality, we find that

$$\left|\frac{\rho_{j+1}}{\rho_j}-\frac{1}{2}\right| \le \frac{1}{2}\frac{c}{\sigma_j}\sin\theta_j+\delta\frac{\rho_j}{{\sigma_j}^2}. \tag{154}$$

Upon setting j to zero and rearranging (154), we obtain

$$\rho_1 \le \rho_0\left(\frac{1}{2}\left(1+\frac{c}{\sigma_0}\right)+\delta\frac{\rho_0}{{\sigma_0}^2}\right). \tag{155}$$

From $x_0 \in \mathcal{R}$ we have $\rho_0 < r(t_0)$ (33), so by the second part of the definition (34), we have

$$\rho_0 < \frac{\sigma_0^2\hat{\rho}(s_0)}{c\sigma_0+\sigma_0^2} = \frac{\hat{\rho}(s_0)}{1+c/\sigma_0},$$

and also

$$\rho_0 < \frac{\sigma_0^2\hat{\rho}(s_0)}{2\delta r_b} \le \frac{\sigma_0^2\hat{\rho}(s_0)}{2\delta\rho_0},$$

where the latter inequality follows from $\rho_0 < r(t_0) \le \bar{r}(t_0) \le r_b$ (see (34), (20)). Applying these inequalities to (155) yields

$$\rho_1 < \frac{1}{2}\hat{\rho}(s_0)+\frac{1}{2}\hat{\rho}(s_0) = \hat{\rho}(s_0).$$

We conclude that if $x_0 \in \mathcal{R}$, then $x_1 \in \mathcal{W}_{s_0}$

A.3 Convergence from $\mathcal{W}_{s_0}$

We next apply the analysis of [8, Section 5], an inductive argument proving linear convergence from inside the domain $\mathcal{W}_{s_)}$.

As in (146), we define $s_0 := g(t_0)/\|g(t_0)\|$. From any initial point $x_1 \in \mathcal{W}_{s_0}$, we show that the sequence of Newton iterates $\{x_j = \rho_j t_j\}_{j\ge 1}$ maintains the properties

$$\rho_j < \hat{\rho}(s_0) \equiv \hat{\rho}, \quad \theta_j < \hat{\phi}(s_0) \equiv \hat{\phi}, \quad \psi_j(s_0) \equiv \psi_j < \phi(s_0) \equiv \phi. \tag{156}$$

By the first and third properties, the iterates remain in $\mathcal{I}_{s_0}$ (29). Further, because of (22), the third property implies that

$$\sigma_j \equiv \sigma(t_j) \ge \hat{\sigma}(s_0) \equiv \hat{\sigma} > 0, \tag{157}$$

a fact that is used often in the proof. We also use the abbreviation

$$q \equiv q(s_0). \tag{158}$$

For $x_1 \in \mathcal{W}_{s_0}$, the first and third properties follow immediately from the previous subsection, as does the second property upon observing that $s \in N$ implies $\sin\theta_j \le \sin\psi_j(s)$.

We assume that (156) holds for all $1 \le i \le j$. From (154) we have for all $i = 1, 2, \ldots, j$ that

$$\begin{aligned}
\left|\frac{\rho_{i+1}}{\rho_i} - \frac{1}{2}\right| &\le \frac{1}{2}\frac{c}{\sigma_i}\sin\hat\phi + \delta\frac{\hat\rho}{{\sigma_i}^2} && \text{from (154) and (156)}\\
&\le \left(\frac{1}{2}\frac{c}{\sigma_i} + \frac{(1-q)\hat\sigma^2}{2{\sigma_i}^2}\right)\sin\hat\phi && \text{from (26) and (157)}\\
&\le \frac{c/\sigma_i + 1 - q}{2}\sin\hat\phi && \text{from (157)}\\
&\le \frac{q}{2}, && \text{from (25).}
\end{aligned}$$

Therefore,

$$\frac{1-q}{2} \le \frac{\rho_{i+1}}{\rho_i} \le \frac{1+q}{2}, \quad \text{for } i = 1, 2, \ldots, j. \tag{159}$$

From the right inequality of (159), we have

$$\rho_{i+1} \le \rho_1\left(\frac{1+q}{2}\right)^i < \hat\rho, \quad \text{for } i = 1, 2, \ldots, j. \tag{160}$$

From (151), we have

(161)

$$\begin{aligned}
\sin\theta_{i+1} &\le \frac{\delta\rho_i^2}{\sigma_i^2\rho_{i+1}} && \text{for } i = 1, 2, \ldots, j\\
&\le \left(\frac{\delta\rho_i}{\sigma_i^2}\right)\left(\frac{2}{1-q}\right) && \text{for } i = 1, 2, \ldots, j, \text{ by the left inequality in (159)}\\
&< \left(\frac{\delta\hat\rho}{\hat\sigma^2}\right)\left(\frac{2}{1-q}\right) && \text{for } i = 1, 2, \ldots, j, \text{ by (156)}\\
&= \sin\hat\phi && \text{for } i = 1, 2, \ldots, j, \text{ by the definition of } \hat\rho \text{ (26).}
\end{aligned}$$

Let $\Delta\psi_j$ be the angle between two consecutive iterates x_j and x_{j+1}. From (153) we have the upper bound

(162)

$$\sin\Delta\psi_i = \min_{\lambda\in\mathsf{R}}\|\lambda x_{i+1} - t_i\| = \frac{2}{\rho_i}\min_{\lambda\in\mathsf{R}}\|\lambda x_{i+1} - \frac{1}{2}x_i\| \le \frac{c}{\sigma_i}\sin\theta_i + \frac{2\delta\rho_i}{\sigma_i^2}, \quad \text{for } i = 1, 2, \ldots, j.$$

In Appendix B, we verify the first equality in (162) by showing that $\Delta\psi_i \le \pi/2$ for $i = 1, 2, \ldots, j$.

By the definition of $\Delta\psi_i$, we have

$$\psi_{j+1} \le \psi_1 + \sum_{i=1}^{j}\Delta\psi_i.$$

Using $\psi_1 < \hat\phi < \pi/2$, $\Delta\psi_i \le \pi/2$, the monotonicity and positivity of sine on $[0, \pi/2]$, and the identity $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha$ for angles α and β, we have

$$\sin\psi_{j+1} \le \sin\hat\phi + \sum_{i=1}^{j}\sin\Delta\psi_i. \tag{163}$$

From (162) and (157), we have

$$\sum_{i=1}^{j} \sin \Delta\psi_i \le \frac{c}{\hat\sigma} \left(\sin\theta_1 + \sum_{i=1}^{j-1} \sin\theta_{i+1} \right) + \frac{2\delta}{\hat\sigma^2} \sum_{i=1}^{j} \rho_i. \tag{164}$$

Using (160), the bound $\rho_1 \le \hat\rho$, and the definition of $\hat\rho$ (26), we have the upper bound

$$\sum_{i=1}^{j} \rho_i \le \rho_1 \frac{2}{1-q} \le \frac{\hat\sigma^2}{\delta} \sin\hat\phi.$$

By combining with bounds like those used in (161), we obtain the upper bound

$$\sum_{i=1}^{j-1} \sin\theta_{i+1} \le \frac{2\delta}{\hat\sigma^2(1-q)} \sum_{i=1}^{j-1} \rho_i \le \frac{2}{(1-q)} \sin\hat\phi.$$

Hence, from (164), we have

$$\sum_{i=1}^{j} \sin \Delta\psi_i \le \frac{c}{\hat\sigma} \left(\sin\theta_1 + \frac{2}{(1-q)} \sin\hat\phi \right) + 2\sin\hat\phi. \tag{165}$$

From (163), by adding $\sin\hat\phi$ to this sum and using $\theta_1 \le \hat\phi$ and the first inequality implicit in the definition of $\sin\hat\phi$ (25), we have

$$\begin{aligned} \sin\psi_{j+1} &< \left(1 + \frac{c}{\hat\sigma}\left(1 + \frac{2}{1-q}\right) + 2\right) \sin\hat\phi \\ &\le \frac{q}{1-q} \left[\frac{(3-q)c/\hat\sigma + (3-3q)}{c/\hat\sigma + 1 - q} \right] \\ &= \frac{q}{1-q} \left[\frac{(3-q)(c/\hat\sigma + 1 - q) + q(1-q)}{c/\hat\sigma + 1 - q} \right] \\ &< \frac{3q}{1-q}. \end{aligned}$$

Since $q = \frac{1}{4}\sin\phi \le \frac{1}{4}$, we find that $\sin\psi_{j+1} < \sin\phi$. From (162) and (166), we have

$$\begin{aligned} \sin\Delta\psi_j &= \frac{2}{\rho_j} \min_{\lambda\in\mathrm{R}} \|\lambda x_{j+1} - \frac{1}{2}x_j\| \le \frac{2}{\rho_j}\|x_{j+1} - \frac{1}{2}x_j\| \\ &\le \frac{2}{\rho_j}\frac{q}{2}\rho_j = q < \sin\phi. \end{aligned}$$

Combining this inequality with $\Delta\psi_j \le \frac{\pi}{2}$, we have $\Delta\psi_j < \phi$. By definition, we must have $\psi_{j+1} \le \psi_j + \Delta\psi_j$. Using our assumption $\psi_j < \phi$ and $\Delta\psi_j < \phi$, we have $\psi_{j+1} < 2\phi \le \frac{\pi}{2}$. Along with $\sin\psi_{j+1} < \sin\phi$, this implies that $\psi_{j+1} < \phi$. Since $\theta_{j+1} \le \psi_{j+1}$ and $\psi_{j+1} \le \frac{\pi}{2}$, we have that (161) implies $\theta_{j+1} < \hat\phi$.

At this point, we have shown that $\rho_{j+1} \le \hat\rho$ (160), $\theta_{j+1} < \hat\phi$ (161), and $\psi_{j+1} < \phi$ (previous paragraph). Hence, (156) continues to hold if j is replaced by $j+1$, and our inductive argument is complete. We conclude that all iterates remain in the set

$$\mathcal{I}_{s_0} := \{x = x^* + \rho t \mid t \in \mathcal{S},\ \cos^{-1}(t^T s_0) < \phi(s_0), 0 < \rho < \hat\rho(s_0)\},$$

which is contained in $\bar{\mathcal{R}}$; see (31).

Noting that (160) and (161) hold for all $j \geq 1$, we see that ρ_j and θ_j go to zero as j goes to infinity. Using these facts in (154), we find that

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

This concludes the proof of Theorem 1, and therefore the extension of Griewank's linear convergence result [8] to Assumption 1.

B Verification of applications of (148)

In this section, we justify the use of the formula (148) by showing that the angle in question is bounded above by $\pi/2$ in each case. As in earlier discussions, we let $\omega_{v,s}$ denote the angle between two vectors $v \in \mathsf{R}^n$, $s \in \mathcal{S}$. We use $\alpha_{v/s}s$ to denote the projection of v onto s, so that $\alpha_{v/s} = v \cdot s = \|v\| \cos \omega_{v,s}$. In each case below, we show that $\alpha_{v/s} \geq 0$, so that $\omega_{v,s} \leq \pi/2$, as desired.

First, we justify the equality in (150). Let $x_{1g} := x_1/\|g_{t_0}\|$, so that $\alpha_{x_{1g}/s_0} s_0$ is the projection of x_{1g} onto s_0. By setting $j = 0$ in (149) and noting that $g(x_0) = \rho_0 g(t_0)$, we have

$$\left\| x_1 - \frac{\rho_0}{2} g(t_0) \right\| \leq \delta \frac{\rho_0^2}{\sigma_0^2}.$$

Dividing by $\|g(t_0)\|$, we have

$$\left\| x_{1g} - \frac{\rho_0}{2} s_0 \right\| \leq \delta \frac{\rho_0^2}{\sigma_0^2 \|g(t_0)\|}.$$

By expressing the vector on the left as a sum of its components parallel to and orthogonal to x_0, we obtain

$$\left\| \alpha_{x_{1g}/s_0} s_0 - \frac{\rho_0}{2} s_0 \right\| \leq \delta \frac{\rho_0^2}{\sigma_0^2 \|g(t_0)\|}.$$

Hence, we have

$$\alpha_{x_{1g}/s_0} \geq \frac{\rho_0}{2} - \delta \frac{\rho_0^2}{\sigma_0^2 \|g(t_0)\|} > \frac{\rho_0}{2} \left(1 - 2\delta \frac{r(t_0)}{\sigma_0^2 \|g(t_0)\|} \right) \geq \frac{\rho_0}{2} \left(1 - \sin \hat{\phi}(s_0) \right) > 0,$$

where the second inequality follows from the $\rho_0 < r(t_0)$, the third inequality follows from the third part of the definition of $r(t)$ (34), and the final (strict) inequality follows from $\hat{\phi} \leq \phi \leq \frac{\pi}{4}$. We conclude that $\alpha_{x_{1g}/s_0} > 0$, as required.

Second, we verify the equality in (162) by showing that $\alpha_{x_{i+1}/t_i} > 0$. By (153), we have for $i \in \{1, 2, \ldots, j\}$,

$$\begin{aligned}
\left\| x_{i+1} - \frac{1}{2} x_i \right\| &\leq \left(\frac{1}{2} \frac{c}{\sigma_i} \sin \theta_i + \delta \frac{\rho_i}{\sigma_i^2} \right) \rho_i \\
&< \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \hat{\phi} + \delta \frac{\hat{\rho}}{\hat{\sigma}^2} \right) \rho_i && \text{by (156)} \\
&= \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + 1 - q \right) \sin \hat{\phi} \rho_i && \text{by (26)} \\
&\leq \frac{q}{2} \rho_i && \text{by (25).}
\end{aligned} \tag{166}$$

By separating into components parallel to and orthogonal to t_i, we have that $\|\alpha_{x_{i+1}/t_i} t_i - \frac{1}{2}\rho_i t_i\| \leq (q/2)\rho_i$, so that $\alpha_{x_{i+1}/t_i} \geq \frac{1}{2}(1-q)\rho_i > 0$ for $i \in \{1, 2, \ldots, j\}$, where the final inequality uses $q < 1$ by (24).

Third, we verify (71) by showing that $\Delta\psi_1 < \pi/2$. From (72),

$$\|x_2 - (1-\alpha/2)x_1\| < \frac{\alpha}{2} q_\alpha \rho_1.$$

Hence, by the usual argument,

$$\|\alpha_{x_2/t_1} t_1 - (1-\alpha/2)\rho_1 t_1\| \leq \frac{\alpha}{2} q_\alpha \rho_1,$$

which implies that

$$\begin{aligned} \alpha_{x_2/t_1} &\geq \left(1 - \alpha/2 - \frac{\alpha}{2} q_\alpha\right)\rho_1 = \left(1 - \alpha/2 - \frac{\alpha}{2}\frac{(1-\alpha/2)}{4}\sin\phi\right)\rho_1 \\ &> (1-\alpha/2)\left(1 - \frac{\sin\phi}{2}\right)\rho_1 > 0 \end{aligned}$$

where we have used the definition of q_α (42) for the equality and the fact that $\alpha < 2$ for the final inequality. Therefore, the angle between x_2 and x_1 must be less than $\pi/2$.

Fourth, we use (91) to justify (98) as follows. From (91), we have by the usual argument that

$$\left\|\alpha_{x_{2k}/t_{2k-2}} t_{2k-2} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)\rho_{2k-2} t_{2k-2}\right\| \leq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) q_\alpha \rho_{2k-2},$$

so that

$$\alpha_{x_{2k}/t_{2k-2}} \geq \frac{1}{2}\left(1 - \frac{\alpha}{2}\right)(1 - q_\alpha)\rho_{2k-2} > 0.$$

Fifth, we justify (102). Since (83) holds for $k = j$, we have

$$\left\|\alpha_{x_{2j+1}/t_{2j}} t_{2j} - \frac{1}{2}\rho_{2j} t_{2j}\right\| < \frac{q_\alpha}{2}\rho_{2j}.$$

Therefore $\alpha_{x_{2j+1}/t_{2j}} \geq \frac{1}{2}(1 - q_\alpha)\rho_{2j} > 0$ as desired.

C 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

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

For illustration, we consider the properties of this problem in some detail. The unique solution is $x^* = (0,1)^T$, where $f(x^*) = (0,0)^T$. Hence $\alpha = \emptyset$, $\beta = \{1\}$, and $\gamma = \{2\}$. We have

$$\Psi'(x^*) = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}.$$

By inspection, we have $N = \{(a, -a)^T,\ a \in \mathsf{R}\}$ and $N_* = \{(b, 0)^T,\ b \in \mathsf{R}\}$. Thus,

$$P_{N_*} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Consider the unit vector $d = \frac{1}{\sqrt{2}}(1, -1)^T$, whose span is N. Using

$$(\Psi')'(x^*; d) = \sqrt{2} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix},$$

we have

$$\bar{B}(d) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} . \sqrt{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}\Bigg|_N = \sqrt{2} \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Consider any $y = (a, -a)^T \in N \setminus \{0\}$. We have $\bar{B}(d)y = \sqrt{2}(-a, a)^T \neq 0$. Thus, $\bar{B}(d)$ is nonsingular, and x^* is a 2^{ae}-regular solution of NCP(f).

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])

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

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

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

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

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

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

Acknowledgements We are grateful to two referees for their very 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 Journal on Optimization **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 Journal on Numerical Analysis **20**, 296–314 (1983)
3. Decker, D.W., Kelley, C.T.: Newton's method at singular points.I. SIAM Journal on Numerical Analysis **17**(1), 66–70 (1980)
4. Decker, D.W., Kelley, C.T.: Convergence acceleration for Newton's method at singular points. SIAM Journal on Numerical Analysis **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 (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. Numerische Mathematik **35**(1), 95–111 (1980)
9. Griewank, A.: On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Review **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 Journal on Numerical Analysis **18**(1), 145–149 (1981)
11. Izmailov, A.F., Solodov, M.V.: Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. Mathematical Programming, Series 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 Journal on Optimization **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. Mathematics of Operations Research **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. Optimization Methods and Software **3**(1), 327–340 (1994)
16. Kelley, C.T., Suresh, R.: A new acceleration method for Newton's method at singular points. SIAM Journal on Numerical Analysis **20**(5), 1001–1009 (1983)
17. ftp://ftp.cs.wisc.edu/math-prog/mcplib
18. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Mathematical Programming **58**, 353–367 (1993)
19. Reddien, G.W.: On Newton's method for singular problems. SIAM Journal on Numerical Analysis **15**(5), 993–996 (1978)
20. Reddien, G.W.: Newton's method and high order singularities. Computers & Mathematics with Applications **5**, 79–86 (1979)