UW Optimization Technical Report 06-02, April 2006

Christina Oberlin · Stephen J. Wright

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

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 dense starlike domain around x^* for F twice Lipschitz continuously differentiable, can be adapted to the case in which F' is only strongly semismooth at the solution. Further, under appropriate regularity assumptions, 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
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

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

where $f : \mathbb{R}^n \to \mathbb{R}^n$. At degenerate solutions of the NCP (for example, solutions x^* at which $x_i^* = f_i(x^*) = 0$ for some i), this nonlinear-equations formulation is not twice differentiable at x^*, so we require the weaker smoothness assumptions considered in this paper to prove convergence of the Newton and accelerated Newton methods. Our results show that (i) the simple approach of applying Newton's method to the nonlinear-equations formulation of the NCP converges from almost any starting point sufficiently close to x^*, albeit at a linear rate if the solution is degenerate; (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) 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 [18], 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.) We focus on a variant of 2-regularity due to Griewank [8], which we refer to it as 2^{ae}-regularity. In the first part of this work, we show that Griewank's analysis, which gives linear convergence from a dense neighborhood of x^*, can be extended to the case in which F' 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. Assuming 2^{ae}-regularity at x^* and F' at least strongly semismooth, we show that this technique yields arbitrarily fast linear convergence from a dense neighborhood of x^*.

In the second part of this work, beginning in Section 6, we consider a particular nonlinear-equations reformulation of the nonlinear complementarity problem (NCP) and interpret the 2^{ae}-regularity condition for this formulation in terms of the properties of the NCP. In particular, we propose a "2^C-regularity condition" for the NCP that implies 2^{ae}-regularity of the nonlinear-equations formulation, and show that this condition reduces to previously known 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 $\Omega \subseteq \mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^m$ we denote the derivative by $G' : \Omega \to \mathbb{R}^m \times \mathbb{R}^n$, that is,

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

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 a 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 $t \in \mathbb{R}^n$ is an *excluded direction* for $\mathcal{A}$ if $\|t\|_2 = 1$ and $x^* + \lambda t \notin \mathcal{A}$ 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 $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $x \in \Omega$, and $d \in \mathbb{R}^n$. If the limit

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

is well defined and exists in $\mathbb{R}^m$, 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]) $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, 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 there exists a neighborhood $\bar{\Omega}$ of $\bar{x}$ such that for all x in $\bar{\Omega}$ different from $\bar{x}$, we have

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

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

Provided F' is strongly semismooth at $\bar{x}$ and $\|x - \bar{x}\|$ is sufficiently small, we have the following crucial estimate from [17]:

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

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 [17]. Hence, there is some $L_{\bar{x}}$ such that

$$\|(F')'(\bar{x}; h_1) - (F')'(\bar{x}; h_2)\| \leq L_{\bar{x}} \|h_1 - h_2\|. \tag{5}$$

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 (for p identified with $n \times n$) 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 $N_* := \ker F'(x^*)^T$. Let $m := \dim N > 0$. We denote by P_N, $P_{N_\perp}$, and P_{N_*} the projection operators into N, $N_\perp$, and N_*, respectively.

The 2-regularity conditions of Reddien [18], Decker and Kelley [3], and Griewank [8] require the following property to hold for certain $r \in N$:

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

In fact, this property appears in the literature as $P_{N_*} F''(x^*)r$ rather than $(P_{N_*} F')'(x^*; r)$. However, our weaker smoothness assumptions in this work require the use of a *directional* second derivative in place of the second derivative tensor operating on a vector. In addition, we follow Izmailov and Solodov [13] in applying P_{N_*} to F' before taking the directional derivative, allowing the theory of 2-regularity to be applied when $P_{N_*} F'$ is directionally differentiable but F' is not.

Decker and Kelley [3] and Reddien [18] use the following definition of 2-regularity.

Definition 4 $2^\forall$-regularity. $2^\forall$-*regularity* holds for F at x^* if (6) holds for every $r \in N \setminus \{0\}$.

Decker and Kelley [4] and Griewank and Osborne [10] observe that this 2-regularity condition is quite strong, since it implies that the dimension of N is at most 2 and the solution x^* is geometrically isolated. We restate the proof of the first of these results which is of particular interest.

Proposition 1 ([4,10]) *If F' is differentiable and* $\dim N > 2$, *then* $2^\forall$-*regularity cannot hold.*

Proof For F' differentiable, $B(r) := (P_{N_*}F')'(x^*;r)|_N$ is a linear function of r. Hence

$$\det(\bar{B}(r)) = (-1)^m \det(\bar{B}(-r)).$$

If $m > 1$, r can be deformed continuously into $-r$ while remaining on the surface of the sphere in N. Thus, if $m > 1$ and m is odd, there must be an $s \in N \cap \mathcal{S}$ with $\det(\bar{B}(s)) = 0$, by the intermediate value theorem for continuous functions. Hence, our claim is proved for $m > 1$ odd. For even m with $m > 2$, we can find a square submatrix of $\bar{B}(r)$ of odd size $\bar{m} > 1$. Applying the above argument to this submatrix, we find that it is singular at some $\bar{s} \in \mathcal{S}$. Hence the whole matrix $\bar{B}$ is singular at $\bar{s}$.

The following weaker 2-regularity, proposed by Griewank [8], can hold regardless of 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 $r \in N$.

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

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

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. 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

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

This matrix is singular only when $d_1 = 0$, that is, it is nonsingular for almost every $d \in \mathbb{R}^2$.

Weaker still is 2^1-regularity.

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

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 is not generally true under the weaker smoothness conditions of this work. However, if $\dim N = 1$, then 2^1-regularity, 2^{ae}-regularity, and $2^\forall$-regularity are all trivially equivalent.

Another 2-regularity condition for $F : \mathbb{R}^n \to \mathbb{R}^n$ was proposed by Izmailov and Solodov [13]. For $h \in \mathbb{R}^n$, consider the linear operator $\Psi_2(h) : \mathbb{R}^n \to \mathbb{R}^n$,

$$\Psi_2(h)\xi := F'(x^*)\xi + (P_{N_*}F')'(x^*;h)\xi$$

and the set

$$T_2 := \{h \in \ker F'(x^*) \mid (P_{N_*}F')'(x^*;h)h = 0\}.$$

Definition 7 $\mathbf{2^T}$-regularity. 2^T*-regularity* holds for $F : \mathbf{R}^n \to \mathbf{R}^n$ at x^* if $T_2 = \{0\}$.

If 2^T-regularity holds, then $\operatorname{range} \Psi_2(h) = \mathbf{R}^n$ for all $h \in \mathbf{R}^n$. Note that 2^T-regularity fails to hold for the example (7).

If $\dim N = 1$, then 2^T-regularity is equivalent to 2^{ae}-regularity, 2^1-regularity, and $2^\forall$-regularity. For completeness, we verify this claim. Suppose $N = \operatorname{span} v$, where $\|v\| = 1$. 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.

3 Prior Work

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

Regularity Conditions. 2-regularity conditions are typically needed to prove convergence of Newton-like methods to singular solutions. As explained in Subsection 2.2, such conditions concern the behavior of certain directional derivatives of F' on the null spaces N and N_* of of $F'(x^*)$ and $F'(x^*)^T$, respectively.

The 2^1-regularity condition (Definition 6) was used by Reddien [19] and Griewank and Osborne [10]. Convergence of Newton's method (at a linear rate of 1/2) can be proved only for starting points x_0 such that $x_0 - x^*$ lies approximately along the particularly direction r along 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. However, as pointed out in [4,10] and in Proposition 1, this condition implies that $\dim N$ is no greater than 2 and that x^* is geometrically isolated. The convergence analysis given for $2^\forall$-regularity [18,3,2] is much simpler than the analysis presented in Griewank [8] and again in this work, which requires weaker regularity assumptions.

Griewank [8, pp. 109-110] shows that when F is sufficiently smooth, the 2^{ae}-regularity condition (Definition 5) is in fact equivalent to 2^1-regularity, and proves convergence of Newton's method from all starting points in a starlike domain with respect to x^* whose set of excluded directions has measure zero—a much more general set than the cones around N analyzed prior to that time.

In the weaker smoothness conditions used in this work, however, the argument used by Griewank to prove equivalence of 2^1- and 2^{ae}-regularity (see proof of Proposition 1) cannot be applied, so we assume explicitly that the 2^{ae}-regularity condition holds.

The 2^{ae}-regularity condition is similar to the 2^T-regularity condition introduced by Izmailov and Solodov for strongly semismooth functions [13], but neither regularity condition appears to imply the other.

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 the null space of $F'(x^*)$. Kelley and Suresh [15] 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 that increases to 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 [15], 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 based on the function

$$\psi_s(a, b) := 2ab - (\min(0, a + b))^2, \tag{8}$$

which has the property that $\psi_s(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$, and $ab = 0$. We define the function $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ by

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

This formulation is apparently due to Evtushenko and Purtov [5] and was studied further by Kanzow [14]. The first derivative Ψ' is strongly semismooth at a solution x^*, if the underlying function f is sufficiently smooth. 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.

In recent years, Izmailov and Solodov [11–13] and Darinya, 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 our 2^C-regularity condition; 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 in general rather than superlinear convergence. Our regularity conditions are comparable and may be weaker in some cases. (For example, the problem munson4 of MCPLIB [16] satisfies both 2^T-regularity and 2^{ae}-regularity but not weak regularity.) We believe that our algorithm has the considerable advantage of simplicity. Near the solution, it modifies Newton's method only by incorporating a simple check to detect linear convergence, and a possible overrelaxation of every second step. The is no need to explicitly 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 [18,3] 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^{ae}-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 smaller and simpler starlike domain $\mathcal{W}_s$, which is similar to the domains of convergence found in earlier works (Reddien [18], Decker and Kelley [3]). Linear convergence is then proved inside $\mathcal{W}_s$. The proof of convergence using $\mathcal{W}_s$ alone is unsatisfactory because it requires not only that the initial iterate is sufficiently close to x^* but also that it lies in a wedge around a certain vector s that lies in the null space N. The use of the larger domain $\mathcal{R}$, which is dense near x^*, removes this limitation to a large extent.

We weaken the smoothness assumption of Griewank and modify 2^{ae}-regularity by replacing the second derivative of F by a directional derivative of F'. Our assumptions follow:

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

We show that Griewank's convergence results hold under our weaker assumptions. Our main result is summarized in the following theorem.

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 $x_0 \in \mathcal{R}$, then 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 (28). Furthermore, for almost every $t_0 \in \mathcal{S}$, there exists a positive number C_{t_0} such that if $\rho_0 < C_{t_0}$ then $x_0 \in \mathcal{R}$.*

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.)

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

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

For $x \in \mathbf{R}^n \setminus \{0\}$, we write $x = \rho t$, where $\rho = \|x\|_2$ and $t = x/\rho$ is a direction in the unit sphere $\mathcal{S}$. By our first assumption in (10), ρ is the 2-norm distance to the solution. 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 and their subscripts indicate their size. 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_*.

Let $(\cdot)|_N$ and $(\cdot)|_{N_\perp}$ denote the restriction maps for N and $N_\perp$. 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

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

and

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

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

$$\begin{aligned} B(x) &= \bar{B}(x) + O(\rho^2) = \rho \bar{B}(t) + O(\rho^2), \\ C(x) &= \bar{C}(x) + O(\rho^2) = \rho \bar{C}(t) + O(\rho^2), \\ D(x) &= O(\rho), \quad \text{and} \quad E(x) = I + O(\rho). \end{aligned} \tag{12}$$

Note that the constants that bound the $O(\cdot)$ terms in these expressions can be chosen independently of t, by Definition 3. This is the first difference between our analysis and Griewank's analysis; we use (4) to arrive at (12), 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 (12), we have

$$G(x) = B(x) + O(\rho^2) = \rho \bar{B}(t) + O(\rho^2). \tag{13}$$

Hence,

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

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

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

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.

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

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

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 (5), $\bar{B}(t)$ is Lipschitz continuous in t, that is, perturbations with respect to t are continuous. Hence, $\sigma(t)$ is continuous in t. This is the second difference between our analysis and Griewank's analysis; we require (5) to prove continuity of $\sigma(t)$, while he applies linearity.

Let

$$\Pi_0(t) := \det \bar{B}(t).$$

It is useful to consider the following restatement of the 2^{ae}-regularity condition: $\Pi_0(t) \neq 0$, for almost every $t \in N$. In contrast to the smooth case considered by Griewank, $\Pi_0(t)$ is not a homogeneous polynomial in t, 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.

4.2 Domains of Invertibility and Convergence

We now define the domains $\mathcal{W}_s$ and $\mathcal{R}$. The following is an adaptation of Lemma 5.1 of [8] to our smoothness and regularity conditions.

Lemma 1 *Suppose Assumption 1 and the standardizations (10) are satisfied. There are two nonnegative continuous functions*

$$\hat{\phi} : N \cap \mathcal{S} \to \mathbf{R} \ \textit{and} \ \hat{\rho} : N \cap \mathcal{S} \to \mathbf{R}$$

such that for any $s \in \mathcal{S} \cap N \setminus {\Pi_0}^{-1}(0)$ the Newton iteration converges linearly with common ratio $1/2$ from all points in the nonempty starlike domain

$$\mathcal{W}_s := \{x = \rho t \,|\, t \in \mathcal{S},\, \cos^{-1}(t^T s) < \hat{\phi}(s), 0 < \rho < \hat{\rho}(s)\}. \tag{15}$$

Further, the iterates remain in the starlike domain

$$\mathcal{I}_s := \{x = \rho t \mid t \in \mathcal{S}, \cos^{-1}(t^T s) < \phi(s), 0 < \rho < \hat{\rho}(s)\}. \tag{16}$$

The definitions of $\hat{\phi}$ and $\hat{\rho}$ depend on several functions that we now introduce. If we define $\min(\emptyset) = \pi$, the angle

$$\phi(s) := \frac{1}{4} \min\{\cos^{-1}(t^T s) \,|\, t \in \mathcal{S} \cap {\Pi_0}^{-1}(0)\}, \quad \text{for } s \in N \cap \mathcal{S} \tag{17}$$

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, $\phi^{-1}(0) = \mathcal{S} \cap N \cap {\Pi_0}^{-1}(0)$ because the set $\{s \in \mathcal{S} \,|\, \Pi_0(s) \neq 0\}$ is open in $\mathcal{S}$ since $\Pi_0(\cdot)$ is continuous on $\mathcal{S}$, by (5).

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

$$\bar{\mathcal{R}} := \{x = \rho t \,|\, t \in \mathcal{S},\, 0 < \rho < \bar{r}(t)\}, \tag{18}$$

where

$$\bar{r}(t) := \min\left\{r_b, \frac{1}{2}\gamma^{-1}\sigma(t)\right\}. \tag{19}$$

Since $\sigma(t)$ is nonzero for almost every $t \in \mathcal{S}$, $\bar{r}(t)$ is nonzero for almost every $t \in \mathcal{S}$.

As in [8, Lemma 5.1], we define

$$\hat{r}(s) := \min\{\bar{r}(t) \,|\, t \in \mathcal{S},\, \cos^{-1}(t^T s) \le \phi(s)\}, \text{ for } s \in N \cap \mathcal{S} \tag{20}$$

and

$$\hat{\sigma}(s) := \min\{\sigma(t) \,|\, t \in \mathcal{S},\, \cos^{-1}(t^T s) \le \phi(s)\}, \text{ for } s \in N \cap \mathcal{S}. \tag{21}$$

These minima exist and both are nonnegative and continuous on $\mathcal{S} \cap N$ with $\hat{\sigma}^{-1}(0) = \hat{r}^{-1}(0) = \phi^{-1}(0)$. For convenience, we reduce $\hat{\sigma}$ as necessary so that

$$\hat{\sigma}(s) \le 1, \text{ for } s \in N \cap \mathcal{S}.$$

Let c be the positive constant defined by

$$c := \max\{\|\bar{C}(t)\| + \sigma(t) \mid t \in \mathcal{S}\}. \tag{22}$$

With the abbreviation

$$q(s) := \frac{1}{4}\sin\phi(s) \le \frac{1}{4}, \text{ for } s \in N \cap \mathcal{S}, \tag{23}$$

we define

$$\sin\hat{\phi}(s) := \min\left\{\frac{q(s)}{c/\hat{\sigma}(s) + 1 - q(s)}, \frac{2\delta\hat{r}(s)}{(1-q(s))\hat{\sigma}^2(s)}\right\}, \text{ for } s \in N \cap \mathcal{S}, \tag{24}$$

where δ is a problem-dependent, positive number to be specified below. By definition, $\sin\hat{\phi}(s) \le \sin\phi(s)$. $\hat{\phi}(s)$ is defined by (24) with the additional requirement that $\hat{\phi}(s) \le \pi/2$. As a result, we have from (24) that $\hat{\phi}(s) \le \phi(s)$, which implies $\mathcal{W}_s \subseteq \mathcal{I}_s$.

We now define

$$\hat{\rho}(s) := \frac{(1-q(s))\hat{\sigma}^2(s)}{2\delta}\sin\hat{\phi}(s), \text{ for } s \in N \cap \mathcal{S}. \tag{25}$$

The second implicit inequality in the definition of $\sin\hat{\phi}(s)$, ensures that $\hat{\rho}(s)$ satisfies

$$\hat{\rho}(s) \le \hat{r}(s) \le \bar{r}(t) \le r_b, \text{ for all } t \in \mathcal{S} \text{ with } \cos^{-1} t^T s \le \phi(s). \tag{26}$$

Both $\hat{\phi}$ and $\hat{\rho}$ are nonnegative and continuous on $N \cap \mathcal{S}$ with $\hat{\phi}^{-1}(0) = \hat{\rho}^{-1}(0) = \phi^{-1}(0)$. The definition of $\mathcal{W}_s$ is now complete, except for the specification of δ, which appears below in (124). For $s \in \mathcal{S} \cap N$, $\mathcal{W}_s = \emptyset$ if and only if $\Pi_0(s) = 0$.

We note that it follows from (26) that

$$\mathcal{I}_s \subset \bar{\mathcal{R}}, \quad \text{for all} \quad s \in \mathcal{S} \cap N \setminus \Pi_0^{-1}(0). \tag{27}$$

(Griewank [8] used the weaker inequality $\hat{r}(s) \leq \bar{r}(s)$ to justify this inclusion, but it is insufficient.)

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

$$g(x) = \rho g(t) = \begin{bmatrix} I & \bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x. \tag{28}$$

The starlike domain of convergence $\mathcal{R}$, which lies inside of the domain $\bar{\mathcal{R}}$, is defined as follows:

$$\mathcal{R} := \{x = \rho t \,|\, t \in \mathcal{S},\, 0 < \rho < r(t)\}, \tag{29}$$

where

$$r(t) := \min\left\{ \bar{r}(t), \frac{\sigma^2(t)\hat{\rho}(s(t))}{2\delta r_b + c\sigma(t) + \sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t)\sin\hat{\phi}(s(t))}{2\delta} \right\}, \tag{30}$$

where we define

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

and δ is a constant to be defined in the next subsection.

(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 (Section A), which picks up the development at this point.

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

$$x_{j+1} = x_j - \alpha F(x_j)^{-1} F(x_j), \tag{31}$$

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, we can show 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 $\frac{1}{3}$. (We omit the precise statement and proof of this result.) 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 (31) for some fixed $\alpha \in (1, 2]$. Broadly speaking, each pure Newton step refocuses the iterates along the null space. Kelley and Suresh prove superlinear convergence for this method when α is systematically increased to 2 as the iterates converge. However, their proof requires the third derivative of F evaluated at x^* to satisfy a boundedness condition and assumes a starting point x_0 that lies near a 2^1-regular direction in the null space of $F'(x^*)$.

We state our main result here and prove it in the remainder of this section. The major assumptions are that 2^{ae}-regularity holds at x^* and that $x_0 \in \mathcal{R}_\alpha$, where $\mathcal{R}_\alpha$ is a starlike domain whose excluded directions are identical to those of $\mathcal{R}$ defined in Section 4 but whose rays may be shorter.

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 $j = 0, 1, 2, \ldots$*

$$x_{2j+1} = x_{2j} - F(x_{2j})^{-1} F(x_{2j}) \quad \textit{and} \tag{32}$$

$$x_{2j+2} = x_{2j+1} - \alpha F(x_{2j+1})^{-1} F(x_{2j+1}), \tag{33}$$

then the iterates converge linearly to x^* with mean rate $\sqrt{\frac{1}{2}\left(1 - \frac{\alpha}{2}\right)}$.

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:

$$\tilde{\delta} \equiv \delta \max(c, \alpha), \tag{34}$$

where δ is defined in (124) and c is defined in (22). Note that

$$\tilde{\delta} \geq \delta. \tag{35}$$

We introduce the following new parameters:

$$q_\alpha(s) := \frac{1 - \alpha/2}{4} \sin \phi(s), \text{ for } s \in N \cap \mathcal{S}, \tag{36}$$

(from which it follows immediately that $q_\alpha(s) \leq (1/8) \sin \phi(s) \leq 1/8$) and

$$\sin \tilde{\phi}_\alpha(s) := \min \left\{ \frac{q_\alpha(s)}{c/\hat{\sigma}(s) + 1 - q_\alpha(s)}, \frac{2\delta \hat{r}(s)}{(1 - q_\alpha(s))\hat{\sigma}^2(s)} \right\}, \text{ for } s \in N \cap \mathcal{S}. \tag{37}$$

The angle $\tilde{\phi}_\alpha(s)$ is defined by (37) with the additional requirement that $\tilde{\phi}_\alpha(s) \leq \pi/2$. As a result, the first part of (37) together with (36) implies that

$$\sin \tilde{\phi}_\alpha(s) \leq \frac{q_\alpha(s)}{1 - q_\alpha(s)} \leq \frac{(1/8) \sin \phi(s)}{1 - (1/8)} \leq \sin \phi(s),$$

and therefore

$$\tilde{\phi}_\alpha(s) \leq \phi(s). \tag{38}$$

We further define

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

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

and

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

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

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

where
(42)
$$r_\alpha(t) := \min\left\{\bar{r}(t), \frac{\sigma^2(t)\tilde{\rho}_\alpha(s(t))}{2\delta r_b + c\sigma(t) + \sigma^2(t)}, \frac{\|g(t)\|\sigma^2(t)(1-\alpha/2)\sin\tilde{\phi}_\alpha(s(t))}{8\delta}\right\}$$

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 $\mathcal{I}_{s,\alpha} \subseteq \mathcal{I}_s \subseteq \bar{\mathcal{R}}$, by demonstrating relationships between the quantities that bound the angles and radii of these sets.

Since $\alpha \ge 1$, a comparison of (23) and (36) yields

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

The definition of $\sin\tilde{\phi}_\alpha$ (37) is simply that of $\sin\hat{\phi}$ (24) with q replaced by q_α. By (43), 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

(44) $$\sin\tilde{\phi}_\alpha(s) \le \sin\hat{\phi}(s).$$

We next show that

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

where $\hat{\rho}(s)$ is defined in (25). Because of (44), it suffices for (45) 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 (35).

Using (44) and (45) together with (35), a comparison of $r_\alpha(t)$ (42) and $r(t)$ (30) yields $r_\alpha(t) \le r(t)$. Therefore, $\mathcal{R}_\alpha \subseteq \mathcal{R}$. The relation $\mathcal{R} \subseteq \bar{\mathcal{R}}$ follows easily from the definitions of r (30) and $\bar{r}$ (19). By (45), we also have $\mathcal{I}_{s,\alpha} \subseteq \mathcal{I}_s$, upon comparing their definitions (41) and (16). The relation $\mathcal{I}_s \subseteq \bar{\mathcal{R}}$ was demonstrated in (27).

As in Section 4, we denote the sequence of iterates by $\{x_i\}_{i\geq 0}$ and use the notation (126), 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 (28). We use the following abbreviations throughout the remainder of this section:

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

5.2 Basic Error Bounds and Outline of Proof

Since the problem is in standard form, we have from (125) that the Newton step (32) satisfies the following relationships for $x_{2k} \in \bar{\mathcal{R}}$:

$$x_{2k+1} = \frac{1}{2}\begin{bmatrix} I & \bar{B}(t_{2k})^{-1}\bar{C}(t_{2k}) \\ 0 & 0 \end{bmatrix} x_{2k} + e(x_{2k}) = \frac{1}{2}g(x_{2k}) + e(x_{2k}), \tag{46}$$

for all $k \geq 0$, where $g(\cdot)$ is defined in (28) and the remainder term $e(\cdot)$ is defined in (123). As in (124), we have

$$\|e(x_{2k})\| \leq \delta \frac{\rho_{2k}^2}{\sigma_{2k}^2}. \tag{47}$$

For the accelerated Newton step (33), we have for $x_{2k+1} \in \bar{\mathcal{R}}$ that

$$x_{2k+2} = \begin{bmatrix} (1-\frac{\alpha}{2})I & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & (1-\alpha)I \end{bmatrix} x_{2k+1} + \alpha e(x_{2k+1}), \tag{48}$$

for all $k \geq 0$, which from (124) yields

$$\begin{aligned} \|x_{2k+2} - (1-\alpha/2)x_{2k+1}\| &\leq \left\| \begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & -\frac{\alpha}{2}I \end{bmatrix} t_{2k+1} \right\| \rho_{2k+1} + \alpha\delta \frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2} \\ &\leq \left\| \begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}(t_{2k+1})^{-1}\bar{C}(t_{2k+1}) \\ 0 & -\frac{\alpha}{2}I \end{bmatrix} t_{2k+1} \right\| \rho_{2k+1} + \tilde{\delta} \frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2}, \end{aligned} \tag{49}$$

where $\tilde{\delta}$ is defined in (34).

By substituting (46) into (48), we obtain

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

where

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

Therefore,

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

$$= \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) g(x_{2k}) + \tilde{e}_\alpha(x_{2k}, x_{2k+1}). \tag{52}$$

To bound the remainder term, we have from (51) that

$$\|\tilde{e}_\alpha(x_{2k}, x_{2k+1})\| \leq \frac{\alpha}{2}\left(1 + \|\bar{B}(t_{2k+1})^{-1}\|\|\bar{C}(t_{2k+1})\|\right)\|e(x_{2k})\| + \alpha\|e(x_{2k+1})\|$$

$$\leq \left(\frac{\sigma_{2k+1} + \|\bar{C}(t_{2k+1})\|}{\sigma_{2k+1}}\right)\delta\frac{\rho_{2k}^2}{\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2}$$

from $\alpha < 2$, (14) and (124)

$$\leq c\delta\frac{\rho_{2k}^2}{\sigma_{2k+1}\sigma_{2k}^2} + \alpha\delta\frac{\rho_{2k+1}^2}{\sigma_{2k+1}^2}$$

from (22)

$$\leq \tilde{\delta}\frac{\rho_{2k}^2 + \rho_{2k+1}^2}{\mu_{2k}^3}, \tag{53}$$

where

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

and $\tilde{\delta}$ is defined as in (34). By combining (52) with (53), we obtain

$$\left\|x_{2k+2} - \frac{1}{2}\left(1 - \frac{\alpha}{2}\right) g(x_{2k})\right\| \leq \tilde{\delta}\frac{\rho_{2k}^2 + \rho_{2k+1}^2}{\mu_{2k}^3}. \tag{55}$$

The proof of Theorem 2 is by induction. The induction step consists of showing that if

(56)

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

then

$$\rho_{2j+\iota} < \tilde{\rho}_\alpha, \quad \theta_{2j+\iota} < \tilde{\phi}_\alpha, \text{ and } \psi_{2j+\iota} < \phi \quad \text{for } \iota \in \{1, 2\} \tag{57}$$

For all $i = 1, 2, \ldots$, the third property in (56), (57)—$\psi_i < \phi$—implies the crucial fact that $\sigma_i \geq \hat{\sigma} > 0$; see (21) and (137). 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 (46) and (48) are valid for our iterates. The convergence rate follows from 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 (56) for $j = 1$, as we now verify. By the definition of (40) (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 (38) and the elementary inequality $\theta_1 \leq \psi_1$, we have $\theta_1 \leq \psi_1 < \tilde{\phi}_\alpha \leq \phi$. Therefore, the inequalities in (56) hold for $k = 0$ and $\iota = 1$. Since $x_2 \in \mathcal{I}_{s_0,\alpha}$, we have $\rho_2 < \tilde{\rho}_\alpha$ and $\psi_2 < \phi$. With the additional fact that $\theta_2 < \tilde{\phi}_\alpha$, we conclude that the inequalities in (56) hold for $k = 0$ and $\iota = 2$. Hence, (56) 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 $\mathcal{R}_\alpha \subseteq \mathcal{R}$, the inequalities of Section 4 remain valid. In particular, we can reuse (135) and write

$$\rho_1 \leq \rho_0 \left(\frac{1}{2} \left(1 + \frac{c}{\sigma_0} \right) + \delta \frac{\rho_0}{\sigma_0^2} \right) \leq \frac{1}{2} \rho_0 \frac{\sigma_0^2 + c\sigma_0 + 2\delta\rho_0}{\sigma_0^2}. \tag{58}$$

Since $x_0 \in \mathcal{R}_\alpha$, we have $\rho_0 < r_\alpha(t_0)$. In addition, since $r_\alpha(t_0) \leq \bar{r}(t_0) \leq r_b$ (which follows from (19) and (42)), we have $\rho_0 < r_b$. Hence, from (58), we have

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

By using the second part of the definition of r_α (42), we thus obtain

$$\rho_1 < \frac{1}{2} \tilde{\rho}_\alpha < \tilde{\rho}_\alpha. \tag{59}$$

As noted above, the inclusion $\mathcal{R}_\alpha \subseteq \mathcal{R}$ implies that inequality (130) is valid here, that is,

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

Since $\rho_0 < r_\alpha(t_0)$, we can apply the third inequality implicit in the definition of r_α (42) to obtain

$$\sin \psi_1(s_0) \leq \frac{1 - \alpha/2}{4} \sin \tilde{\phi}_\alpha(s_0) < \sin \tilde{\phi}_\alpha(s_0). \tag{60}$$

We have now shown that $x_1 \in \mathcal{W}_{s_0,\alpha}$. We note that (60) and (38) imply that $\psi_1 < \tilde{\phi}_\alpha \leq \phi$, which implies $\sigma_1 \geq \hat{\sigma}$, by the definition of $\hat{\sigma}$ (21).

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 (59). From (48) for $j = 0$, we have by decomposing x_1 into components in N and $N_\perp$ that

$$\begin{aligned}
\rho_2 &\le \left(\left(1-\frac{\alpha}{2}\right)\cos\theta_1 + \left(\frac{\alpha\|\bar{C}(t_1)\|}{2\sigma_1} + \alpha - 1\right)\sin\theta_1 + \alpha\delta\frac{\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\qquad \text{from (14), (47), and (124)} \\
&\le \left(1-\frac{\alpha}{2} + \left(\frac{\alpha}{2}\frac{\|\bar{C}(t_1)\|+\sigma_1}{\sigma_1} + \frac{\alpha}{2} - 1\right)\sin\theta_1 + \tilde{\delta}\frac{\rho_1}{\sigma_1^2}\right)\rho_1 \\
&\qquad \text{from } \cos\theta_1 \le 1 \text{ and (34)} \\
&\le \left(1-\frac{\alpha}{2} + \left(\frac{\alpha}{2}\frac{c}{\hat{\sigma}} + \frac{\alpha}{2} - 1\right)\sin\theta_1 + \tilde{\delta}\frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2}\right)\rho_1 \\
&\qquad \text{from (22), } \sigma_1 \ge \hat{\sigma}, \text{ and } \rho_1 < \tilde{\rho}_\alpha \\
&< \left(1-\frac{\alpha}{2} + \frac{\alpha}{2}\left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha\right)\sin\tilde{\phi}_\alpha + \tilde{\delta}\frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2}\right)\rho_1 \\
&\qquad \text{from } \alpha \in [1,2),\ q_\alpha < \tfrac{1}{8}, \text{ and } \theta_1 < \tilde{\phi}_\alpha.
\end{aligned}$$

By replacing $\sin\tilde{\phi}_\alpha$ with the first inequality implicit in its definition (37) and using the definition of $\tilde{\rho}_\alpha$ (39), 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 (37), the definition of c (22), and $q_\alpha < \frac{1}{8}$, we have

$$\sin\tilde{\phi}_\alpha < \frac{q_\alpha}{2-(1/8)} < q_\alpha. \tag{61}$$

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 (60), (38), and (17), we have $\psi_1 \le \tilde{\phi}_\alpha \le \phi \le \pi/4$. In Appendix B, we demonstrate that $\Delta\psi_1 \le \pi/2$. Thus, using (60), we have

$$\sin\psi_2 \le \sin\psi_1 + \sin\Delta\psi_1 \le \frac{1-\alpha/2}{4}\sin\tilde{\phi}_\alpha + \sin\Delta\psi_1. \tag{62}$$

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

$$\sin\Delta\psi_1 \equiv \min_{\lambda\in\mathbf{R}}\|\lambda x_2 - t_1\|. \tag{63}$$

By (49) with $j = 0$, we have

(64) $$\begin{aligned} &\|x_2 - (1-\alpha/2)x_1\| \\ &\le \left(\left\| \begin{bmatrix} 0 & \frac{\alpha}{2}\bar{B}^{-1}(t_1)\bar{C}(t_1) \\ 0 & \frac{-\alpha}{2} I \end{bmatrix} t_1 \right\| + \frac{\tilde{\delta}\rho_1}{\sigma_1^2} \right) \rho_1 \\ &\le \left(\frac{\alpha}{2} \frac{c}{\sigma_1} \sin\theta_1 + \frac{\tilde{\delta}\rho_1}{\sigma_1^2} \right) \rho_1 \\ &\qquad \text{by (14) and (22)} \\ &< \frac{\alpha}{2} \left(\frac{c}{\hat{\sigma}} \sin\tilde{\phi}_\alpha + 2\frac{\tilde{\delta}\tilde{\rho}_\alpha}{\hat{\sigma}^2} \right) \rho_1 \\ &\qquad \text{by } \sigma_1 \ge \hat{\sigma},\ \sin\theta_1 < \sin\tilde{\phi}_\alpha,\ \rho_1 < \tilde{\rho}_\alpha(s), \text{ and } \alpha \ge 1 \\ &= \frac{\alpha}{2} \left(\frac{c}{\hat{\sigma}} \sin\tilde{\phi}_\alpha + \frac{(1-\alpha/2-q_\alpha)\hat{\sigma}}{2} \sin\tilde{\phi}_\alpha \right) \rho_1 \\ &\qquad \text{by (39)} \\ &\le \frac{\alpha}{2} \left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha \right) \sin\tilde{\phi}_\alpha \rho_1 \\ &\qquad \text{by } \alpha \ge 0,\ q \le 1, \text{ and } \hat{\sigma} \le 1 \\ &\le \frac{\alpha}{2} q_\alpha \rho_1 \\ &\qquad \text{by the first inequality in (37).} \end{aligned}$$

The inequality (64) provides a bound on $\sin \varDelta\psi_1$ in terms of $q_\alpha(s)$:

(65) $$\sin \varDelta\psi_1 = \min_{\lambda \in \mathbf{R}} \|\lambda x_2 - t_1\| \le \left\| \frac{x_2}{(1-\alpha/2)\rho_1} - t_1 \right\| < \frac{\alpha}{2(1-\alpha/2)} q_\alpha.$$

By substituting (65) into (62) and using (36), we find that

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

From (61), (36), and $(1-\alpha/2) \in (0, 1/2]$, we have

(67) $$\sin\tilde{\phi}_\alpha < \frac{8}{15} q_\alpha \le \frac{1}{15} \sin\phi.$$

Therefore, we have

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

To complete the anchor argument, we need to show that $\sin\theta_2 < \sin\tilde{\phi}_\alpha$. From the second row of (48) with $j = 0$, and using (124) with $x = x_1$ and (34), we have

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

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 (60) and the definition of $\tilde{\rho}_\alpha$ (39), we have

$$\begin{aligned}(69)\qquad \rho_2 \sin\theta_2 &< \rho_1 \left((\alpha - 1)\frac{(1-\alpha/2)}{4} \sin\tilde{\phi}_\alpha + \frac{(1-\alpha/2-q_\alpha)\hat{\sigma}}{4} \sin\tilde{\phi}_\alpha \right) \\ &\leq \rho_1 \left(\alpha \frac{(1-\alpha/2)}{4} \sin\tilde{\phi}_\alpha \right) < \rho_1 \frac{(1-\alpha/2)}{2} \sin\tilde{\phi}_\alpha,\end{aligned}$$

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 (69), we require a lower bound on ρ_2 in terms of a fraction of ρ_1. By applying the inverse triangle inequality to (64), 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 (36), 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 (69), we find that

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

Hence, $\sin\theta_2 < \sin\tilde{\phi}_\alpha$, as desired.

5.4 The Induction Step

In the remainder of this proof, we provide the argument for the induction step: If (56) holds for some j, then (57) holds as well.

5.4.1 Iteration $2j + 1$

We show in this subsection that if (56) holds, that is,

$$(70)\qquad \rho_i < \tilde{\rho}_\alpha, \quad \sin\theta_i < \sin\tilde{\phi}_\alpha, \quad \sin\psi_i < \sin\phi, \quad \text{for } i = 1, 2, \ldots, 2j,$$

then after the step from x_{2j} to x_{2j+1}, which is a regular Newton step, we have (57) for $\iota = 1$, that is,

$$(71)\qquad \rho_{2j+1} < \tilde{\rho}_\alpha, \quad \sin\theta_{2j+1} < \sin\tilde{\phi}_\alpha, \quad \sin\psi_{2j+1} < \sin\phi.$$

Consider $k \in \{1, 2, \ldots, j\}$. In the same manner that the inequalities (131) and (133) follow from equation (129), we have the following nearly identical inequalities (72) and (73) following from equations (46) and (47):

$$(72)\qquad \sin\theta_{2k+1} = \min_{y \in N} \|t_{2k+1} - y\| \leq \delta \frac{\rho_{2k}^2}{\sigma_{2k}^2 \rho_{2k+1}}$$

and

$$\left\| x_{2k+1} - \frac{1}{2} x_{2k} \right\| \leq \left(\frac{1}{2} \frac{c}{\sigma_{2k}} \sin \theta_{2k} + \delta \frac{\rho_{2k}}{\sigma_{2k}^2} \right) \rho_{2k}. \tag{73}$$

By dividing (73) by ρ_{2k} and applying the reverse triangle inequality, we have

$$\left| \frac{\rho_{2k+1}}{\rho_{2k}} - \frac{1}{2} \right| \leq \left(\frac{1}{2} \frac{c}{\sigma_{2k}} \sin \theta_{2k} + \delta \frac{\rho_{2k}}{\sigma_{2k}^2} \right). \tag{74}$$

Further, from (56), we can bound (73) as follows, for all $k = 1, 2, \ldots, j$:

(75)

$$\begin{aligned}
&\left\| x_{2k+1} - \frac{1}{2} x_{2k} \right\| \\
&< \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \tilde{\phi}_\alpha + \delta \frac{\tilde{\rho}_\alpha}{\hat{\sigma}^2} \right) \rho_{2k} && \text{using } \sigma_{2k} \geq \hat{\sigma},\ \theta_{2k} < \tilde{\phi}_\alpha,\ \rho_{2k} < \tilde{\rho}_\alpha \\
&\leq \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \tilde{\phi}_\alpha + \frac{(1 - \alpha/2 - q_\alpha)\hat{\sigma}}{4} \sin \tilde{\phi}_\alpha \right) \rho_{2k} && \text{by (35) and (39)} \\
&< \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + \frac{(1 - \alpha/2)}{2} \right) \rho_{2k} \sin \tilde{\phi}_\alpha && \text{using } q_\alpha > 0 \text{ and } \hat{\sigma} \leq 1 \\
&< \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + 1 - q_\alpha \right) \rho_{2k} \sin \tilde{\phi}_\alpha && \text{using } q_\alpha < \tfrac{1}{8} \text{ and } \alpha \geq 1 \\
&\leq \frac{q_\alpha}{2} \rho_{2k} && \text{by the first part of (37).}
\end{aligned}$$

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

$$\left| \frac{\rho_{2k+1}}{\rho_{2k}} - \frac{1}{2} \right| < \frac{q_\alpha}{2}. \tag{76}$$

Therefore,

$$\frac{1 - q_\alpha}{2} \leq \frac{\rho_{2k+1}}{\rho_{2k}} \leq \frac{1 + q_\alpha}{2}. \tag{77}$$

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

$$\rho_{2k+1} < \rho_{2k} < \tilde{\rho}_\alpha. \tag{78}$$

In particular, we have $\rho_{2j+1} < \tilde{\rho}_\alpha$. From the left inequality and (72), we have

$$\begin{aligned}
\sin \theta_{2k+1} &\leq \delta \frac{2\rho_{2k}}{\sigma_{2k}^2 (1 - q_\alpha)} \\
&< \delta \frac{2\tilde{\rho}_\alpha}{\hat{\sigma}^2 (1 - q_\alpha)} && \text{using } \rho_{2k} < \tilde{\rho}_\alpha \text{ and } \sigma_{2k} \geq \hat{\sigma} \\
&= \frac{(1 - \alpha/2 - q_\alpha)\hat{\sigma}}{2(1 - q_\alpha)} \sin \tilde{\phi}_\alpha && \text{by (35) and (39)} \\
&< \sin \tilde{\phi}_\alpha && \text{using } \hat{\sigma} \leq 1,
\end{aligned} \tag{79}$$

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 (70) and (21) that

$$\sigma_i \geq \hat{\sigma}, \quad i = 1, 2, \ldots, 2j, \tag{80}$$

so it follows from the definition (54) that

$$\mu_{2k-2} \geq \hat{\sigma}, \quad k = 2, 3, \ldots, j. \tag{81}$$

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

$$\begin{aligned} \sin\theta_{2k} &= \min_{y \in N} \|t_{2k} - y\| \\ &\leq \tilde{\delta}\frac{\rho_{2k-2}^2 + \rho_{2k-1}^2}{\mu_{2k-2}^3 \rho_{2k}} && \text{from (55), with } k \leftarrow k-1 \\ &\leq 2\tilde{\delta}\frac{\rho_{2k-2}^2}{\mu_{2k-2}^3 \rho_{2k}} && \text{from (78).} \end{aligned} \tag{82}$$

From (132), 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 (55) (with $k \leftarrow k-1$), we obtain

$$
\begin{aligned}
(83)\quad & \left\| x_{2k} - \frac{1}{2}\left(1-\frac{\alpha}{2}\right) x_{2k-2} \right\| \\
& \le \left\| x_{2k} - \frac{1}{2}\left(1-\frac{\alpha}{2}\right) g(x_{2k-2}) \right\| + \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \| x_{2k-2} - g(x_{2k-2}) \| \\
& \le \tilde{\delta} \frac{\rho_{2k-2}^2 + \rho_{2k-1}^2}{\mu_{2k-2}^3} + \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \frac{c}{\sigma_{2k-2}} \rho_{2k-2} \sin\theta_{2k-2} \\
& \le 2\tilde{\delta} \frac{\rho_{2k-2}^2}{\mu_{2k-2}^3} + \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \frac{c}{\sigma_{2k-2}} \rho_{2k-2} \sin\theta_{2k-2} \\
& \qquad \text{from (78)} \\
& \le \left[2\tilde{\delta} \frac{\tilde{\rho}_\alpha}{\hat{\sigma}^3} + \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \frac{c}{\hat{\sigma}} \sin\tilde{\phi}_\alpha \right] \rho_{2k-2} \\
& \qquad \text{from (70), (81), and (80)} \\
& = \left[\frac{1}{2}\left(1-\frac{\alpha}{2} - q_\alpha\right) + \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
& \qquad \text{from (39)} \\
& \le \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \left[1 - \frac{q_\alpha}{1-\alpha/2} + \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
& \le \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \left[1 - q_\alpha + \frac{c}{\hat{\sigma}} \right] \sin\tilde{\phi}_\alpha \rho_{2k-2} \\
& \qquad \text{from } 0 < 1-\alpha/2 < 1 \text{ and } q_\alpha > 0 \\
& \le \frac{1}{2}\left(1-\frac{\alpha}{2}\right) q_\alpha \rho_{2k-2} \qquad \text{from (37).}
\end{aligned}
$$

Upon dividing by ρ_{2k-2} and applying the reverse triangle inequality, we find from the fourth line of (83) that

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

while from the last line of (83), we have

$$
(85) \qquad \left| \frac{\rho_{2k}}{\rho_{2k-2}} - \frac{1}{2}\left(1-\frac{\alpha}{2}\right) \right| \le \frac{1}{2}\left(1-\frac{\alpha}{2}\right) q_\alpha.
$$

We can restate this inequality as follows:

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

From the right inequality in (86), we obtain

$$
(87) \qquad \rho_{2k} \le \left[\frac{1}{2}\left(1-\frac{\alpha}{2}\right)(1+q_\alpha) \right]^{k-1} \rho_2,
$$

while by substituting the left inequality into (82) and using (81), we obtain

$$
\begin{aligned}
\sin\theta_{2k} &\le 2\tilde{\delta}\frac{\rho_{2k-2}^2}{\mu_{2k-2}^3\rho_{2k}} \\
&\le 4\tilde{\delta}\frac{\rho_{2k-2}}{\hat{\sigma}^3}\frac{1}{(1-\alpha/2)(1-q_\alpha)} \\
&\le 4\tilde{\delta}\frac{\rho_{2k-2}}{\hat{\sigma}^3}\frac{1}{1-\alpha/2-q_\alpha} \qquad \text{for } k=2,3,\ldots,j.
\end{aligned}
\tag{88}
$$

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

$$
\psi_{2j+1} \le \psi_2 + \sum_{k=2}^{j}\Delta^2\psi_{2k-2} + \Delta\psi_{2j}.
\tag{89}
$$

From the fourth line of (83), we have

$$
\begin{aligned}
\sin\Delta^2\psi_{2k-2} &= \min_{\lambda\in\mathbf{R}}\|\lambda x_{2k} - t_{2k-2}\| \\
&= \frac{2}{(1-\alpha/2)\rho_{2k-2}}\min_{\lambda\in\mathbf{R}}\left\|\lambda x_{2k} - \frac{1}{2}\left(1-\frac{\alpha}{2}\right)x_{2k-2}\right\| \\
&\le \frac{4\tilde{\delta}}{(1-\alpha/2)}\frac{\rho_{2k-2}}{\mu_{2k-2}^3} + \frac{c}{\sigma_{2k-2}}\sin\theta_{2k-2} \\
&\le \frac{4\tilde{\delta}\rho_{2k-2}}{(1-\alpha/2)\hat{\sigma}^3} + \frac{c}{\hat{\sigma}}\sin\theta_{2k-2},
\end{aligned}
\tag{90}
$$

by (80) and (81). 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 (90). From (87), we have

$$
\begin{aligned}
\sum_{k=2}^{j}\rho_{2k-2} &\le \sum_{k=0}^{j-2}\left[\frac{1}{2}\left(1-\frac{\alpha}{2}\right)(1+q_\alpha)\right]^k\rho_2 \\
&\le \left[1-\frac{1}{2}\left(1-\frac{\alpha}{2}\right)(1+q_\alpha)\right]^{-1}\rho_2 \\
&= \left[\frac{1}{2}-\frac{1}{2}q_\alpha+\frac{\alpha}{4}+\frac{\alpha}{4}q_\alpha\right]^{-1}\rho_2 \\
&\le \left[\frac{1}{2}-\frac{1}{2}q_\alpha+\frac{\alpha}{4}\right]^{-1}\rho_2 \\
&< \frac{2}{1+(\alpha/2)-q_\alpha}\tilde{\rho}_\alpha \qquad \text{from (70)} \\
&= \frac{1}{2\tilde{\delta}}\frac{1-(\alpha/2)-q_\alpha}{1+(\alpha/2)-q_\alpha}\hat{\sigma}^3\sin\tilde{\phi}_\alpha \qquad \text{from (39).}
\end{aligned}
\tag{91}
$$

From (88) and (91) we have

(92)
$$\begin{aligned}\sum_{k=2}^{j}\sin\theta_{2k-2} &\le \sin\theta_2 + \sum_{k=2}^{j-1}\sin\theta_{2k}\\ &\le \sin\tilde{\phi}_\alpha + \frac{4\tilde{\delta}}{\hat{\sigma}^3}\frac{1}{1-(\alpha/2)-q_\alpha}\sum_{k=2}^{j-1}\rho_{2k-2}\\ &\le \sin\tilde{\phi}_\alpha + \frac{2}{1+(\alpha/2)-q_\alpha}\sin\tilde{\phi}_\alpha\\ &\le \sin\tilde{\phi}_\alpha + \frac{2}{11/8}\sin\tilde{\phi}_\alpha \qquad \text{since } 1+\tfrac{\alpha}{2}-q_\alpha \ge 1+\tfrac{1}{2}-\tfrac{1}{8}=\tfrac{11}{8}\\ &= \frac{27}{11}\sin\tilde{\phi}_\alpha.\end{aligned}$$

By summing (90) over $k = 2, 3, \ldots, j$ and using (91) and (92), we obtain

$$\begin{aligned}(93) \qquad & \sum_{k=2}^{j}\sin\Delta^2\psi_{2k-2}\\ &\le \frac{4\tilde{\delta}}{(1-\alpha/2)\hat{\sigma}^3}\frac{\hat{\sigma}^3}{2\tilde{\delta}}\frac{1-(\alpha/2)-q_\alpha}{1+(\alpha/2)-q_\alpha}\sin\tilde{\phi}_\alpha + \frac{c}{\hat{\sigma}}\frac{27}{11}\sin\tilde{\phi}_\alpha\\ &\le \left[\frac{2}{1+(\alpha/2)-q_\alpha} + \frac{27}{11}\frac{c}{\hat{\sigma}}\right]\sin\tilde{\phi}_\alpha\\ &\le \left[\frac{16}{11} + \frac{27}{11}\frac{c}{\hat{\sigma}}\right]\sin\tilde{\phi}_\alpha,\end{aligned}$$

where we used $q_\alpha > 0$ for the second-last inequality and $1+(\alpha/2)-q_\alpha \ge 11/8$ for the final inequality. For a bound on the term $\Delta\psi_{2j}$ of (89), we use (75) to obtain

$$\begin{aligned}(94) \qquad \sin\Delta\psi_{2j} &= \min_{\lambda\in\mathbf{R}}\|t_{2j}-\lambda x_{2j+1}\|\\ &= \frac{2}{\rho_{2j}}\min_{\lambda\in\mathbf{R}}\left\|\frac{1}{2}x_{2j}-\lambda x_{2j+1}\right\|\\ &\le \frac{2}{\rho_{2j}}\left\|\frac{1}{2}x_{2j}-x_{2j+1}\right\|\\ &\le \left(\frac{c}{\hat{\sigma}}+1-q_\alpha\right)\sin\tilde{\phi}_\alpha.\end{aligned}$$

The equality in (94) follows from the fact that $\Delta\psi_{2j} \le \pi/2$, as shown in Appendix B.

Since each of the angles in (89) is bounded above by $\pi/2$, from reasoning similar to that of Section A.3, we have

$$(95) \qquad \sin\psi_{2j+1} \le \sin\psi_2 + \sum_{k=2}^{j}\sin\Delta^2\psi_{2k-2} + \sin\Delta\psi_{2j}.$$

By substituting (68), (93), and (94) into (95), we obtain

(96)

$$
\begin{aligned}
\sin\psi_{2j+1} &\le \left[\frac{1}{120}+\frac{\alpha}{8}\right]\sin\phi + \left[\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}\right]\sin\tilde\phi_\alpha + \left[\frac{c}{\hat\sigma}+1-q_\alpha\right]\sin\tilde\phi_\alpha \\
&\le \frac{2}{7}\sin\phi + \frac{\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}}{1-q_\alpha+\frac{c}{\hat\sigma}} q_\alpha + q_\alpha \\
&\qquad\qquad \text{from (37) and } \alpha < 2 \\
&\le \frac{2}{7}\sin\phi + \frac{\frac{16}{11}+\frac{27}{11}\frac{c}{\hat\sigma}}{\frac{7}{8}+\frac{c}{\hat\sigma}} q_\alpha + q_\alpha \\
&\qquad\qquad \text{from } q_\alpha < \tfrac{1}{8} \\
&\le \frac{2}{7}\sin\phi + \left(\frac{27}{11}+1\right) q_\alpha \\
&\le \frac{2}{7}\sin\phi + \frac{38}{11}\frac{1}{8}\sin\phi \\
&< \sin\phi.
\end{aligned}
$$

Hence, $\sin\psi_{2j+1} < \sin\phi$, as required.

5.4.2 Iteration $2j+2$

We now show that

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

by using some of the bounds proved above: $\rho_{2j} < \tilde\rho_\alpha$, $\rho_{2j+1} < \rho_{2j}$, $\sin\theta_{2j} < \sin\tilde\phi_\alpha$, $\sin\psi_{2j} < \sin\phi$, and $\sin\psi_{2j+1} < \sin\phi$. The last two assumptions guarantee that $\mu_{2j} \ge \hat\sigma$. The analysis in the latter part of Subsection 5.4.1 (starting from (81)) can therefore be applied for $k = j+1$. In particular, from (87) we have $\rho_{2j+2} < \tilde\rho_\alpha$. From (88) with $k = j+1$, using $\rho_{2j} < \tilde\rho_\alpha$, $\mu_{2j} \ge \hat\sigma$, and the definition of $\tilde\rho_\alpha$ (39) we have

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

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 (95) to $j+1$ (ignoring the final nonnegative term) to give

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

The bounds (87), (88), and (90) continue to hold for $k = j+1$, while (91) and (92) continue to hold if the upper bound on the summation is increased from j to $j+1$, so (93) also continues to hold if the upper bound of the summation is increased from j to $j+1$. Hence, similarly to (96), we obtain $\sin\psi_{2j+2} < \sin\phi$, as required.

Our proof of the induction step is complete.

5.5 Convergence Rate

As $j \to \infty$, we have $\rho_{2j} \to 0$ by (87) and $\rho_{2j+1} \to 0$ by (78). From (88), we also have $\theta_{2j} \to 0$. By combining these limits with (80) and (81), we see that the right-hand-side of (84) goes to zero as $j \to \infty$, and

$$\lim_{j\to\infty} \frac{\rho_{2j+2}}{\rho_{2j}} = \frac{1}{2}\left(1 - \frac{\alpha}{2}\right). \tag{98}$$

From the discussion above and (74), we also have $\lim_{j\to\infty} \frac{\rho_{2j+1}}{\rho_{2j}} = 1/2$, so that convergence is stable between the accelerated iterates. By (98), the mean rate of convergence is $\sqrt{1/2(1-\alpha/2)}$.

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

$$0 \le f(x), \quad x \ge 0, \quad x^T f(x) = 0.$$

We assume that f is twice Lipschitz continuously differentiable.

6.1 NCP Notation, Definitions, and Properties

For any matrix $M \in \mathbb{R}^p \times \mathbb{R}^q$ and any sets $\mathcal{S} \subseteq \{1, 2, \ldots, p\}$ and $\mathcal{T} \subseteq \{1, 2, \ldots, q\}$, we write $M_{\mathcal{S},\mathcal{T}}$ to denote the submatrix of M whose rows lie in $\mathcal{S}$ and columns lie in $\mathcal{T}$. We denote the number of elements in any set $\mathcal{S}$ by $|\mathcal{S}|$. 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 $\operatorname{diag} x$ denote the $\mathbb{R}^{n\times n}$ diagonal matrix formed from the components of x.

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

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

We state the following definitions for later reference.

Definition 8 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.

6.2 The Nonlinear-Equations Reformulation

We recall the nonlinear-equations reformulation (9) of the NCP (1), and consider the use of Newton's method for solving $\Psi(x) = 0$. In particular, we tailor the local convergence result of Section 4, which is for the general function F, to the function Ψ defined above. In this section, we establish the structure of null space N and the form of the 2^{ae}-regularity condition for Ψ, then rewrite Theorem 1 for Ψ.

Taking the derivative of Ψ, we have

$$\begin{aligned}\Psi_i'(x) = 2\{&(f_i(x) - \min(0, x_i + f_i(x)))e_i \\ &+ (x_i - \min(0, x_i + f_i(x)))f_i'(x)\}, \quad \text{for } i = 1, 2, \ldots, n.\end{aligned} \tag{99}$$

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

$$N \equiv \ker \Psi'(x^*) = \{\xi \in \mathbb{R}^n \mid f_\gamma'(x^*)\xi = 0,\ \xi_\alpha = 0\}, \tag{100}$$

so that

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

In particular, if $\beta \neq \emptyset$, then $\dim N > 1$ and x^* is a singular solution of $\Psi(x) = 0$. The null space of $\Psi'(x^*)^T$ is

$$\begin{aligned} N_* = \{\xi \in \mathbb{R}^n \mid \xi_\alpha = -(\operatorname{diag} f_\alpha(x^*))^{-1}(f_{\gamma,\alpha}'(x^*))^T(\operatorname{diag} x_\gamma^*)\xi_\gamma, \\ f_{\gamma,\beta\cup\gamma}'(x^*)^T(\operatorname{diag} x_\gamma^*)\xi_\gamma = 0\}. \end{aligned} \tag{101}$$

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

We now examine the 2^{ae}-regularity condition for Ψ at x^*. By direct calculation, we have

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

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

$$(\Psi_i')'(x^*; d) = 2 \begin{cases} \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*), & i \in \alpha, \\ (\langle f_i'(x^*), d\rangle - \min(0, d_i + \langle f_i'(x^*), d\rangle))e_i & \\ \quad + (d_i - \min(0, d_i + \langle f_i'(x^*), d\rangle)) f_i'(x^*), & i \in \beta, \\ \langle f_i'(x^*), d\rangle e_i + d_i f_i'(x^*) + x_i^* (f_i')'(x^*; d), & i \in \gamma \end{cases} \tag{102}$$

A solution x^* is 2^{ae}-regular if $\bar{B}(d) := (P_{N_*}\Psi')'(x^*; d)|_N$ is nonsingular for almost every $d \in N \backslash \{0\}$. Such values of d allow for the following simplification of (102):

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

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 the expression above as follows, for $d \in N$:

$$\frac{1}{2}(\Psi_i')'(x^*; d) = \begin{cases} \langle f_i'(x^*), d\rangle e_i, & i \in \alpha, \\ \max(\langle f_i'(x^*), d\rangle, -d_i)e_i - \min(\langle f_i'(x^*), d\rangle, -d_i) f_i'(x^*), & i \in \beta, \\ d_i f_i'(x^*) + x_i^* (f_i')'(x^*; d), & i \in \gamma \end{cases} \tag{103}$$

From [6, Proposition 7.4.4], the composition of strongly semismooth functions is strongly semismooth. Hence, it follows from the definition (9) and the properties of ψ_s noted above that Ψ' is strongly semismooth when f' is strongly semismooth. (In particular, when f is twice Lipschitz continuously differentiable, then Ψ' is strongly semismooth.)

The following theorem specializes Theorem 1 for applying Newton's method to $\Psi(x)$.

Theorem 3 *Consider the 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 2^{ae}-regularity holds at x^* for the the function Ψ defined in (9), that is,*

$$(P_{N_*}\Psi')'(x^*; d)|_N \text{ is nonsingular for almost every } d \in N \setminus \{0\}. \tag{104}$$

Then there exists a starlike domain $\mathcal{R}$ about x^*, which is dense near x^*, such that, if Newton's method for $\Psi(x)$ 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 (32) and (33) for some $\alpha \in [1, 2)$, then there exists a starlike domain $\mathcal{R}_\alpha \subseteq \mathcal{R}$ about x^*, which is dense near x^*, such that if $x_0 \in \mathcal{R}_\alpha$ then the accelerated iterates converge linearly to x^* with mean rate $\sqrt{\frac{1}{2}\left(1 - \frac{\alpha}{2}\right)}$.

We elucidate condition (104) by considering several special cases of the NCP(f) for $f : \mathbb{R}^n \to \mathbb{R}^n$ and solution $x^* \in \mathbb{R}^n$.

6.3 Regularity Conditions for NCP

In this section we examine the regularity condition (104) and find conditions on the NCP under which it is satisfied. In particular, we introduce a regularity condition for NCP called 2^C*-regularity*, prove that it implies (104), and show how it relates to more familiar, previously proposed regularity conditions for NCP in certain special cases.

As a preliminary, we introduce some notation. First, we let

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

We then 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 W of dimensions $|\gamma| \times (|\gamma| - r)$ such that the columns of W span $\ker f'_{\gamma,\beta\cup\gamma}(x^*)^T$. Note that $[Z \,|\, W]$ is an orthogonal matrix of dimensions $|\gamma| \times |\gamma|$.

Definition 9 If for almost every $d \in N \setminus \{0\}$ the following $n \times n$ matrix is nonsingular:

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

we say that 2^C*-regularity* holds for NCP at x^*.

Proposition 2 *If the NCP satisfies the 2^C-regularity condition at the solution x^*, then the function Ψ satisfies the 2^{ae}-regularity condition (104) at x^*.*

Proof We drop the argument x^* from f, f', and so on, for clarity.

To prove the desired result, we need to show that for almost all $d \in N\setminus\{0\}$, we have that

$$P_{N_*}(\Psi')'(x^*; d)v = 0 \text{ and } v \in N \ \Rightarrow\ v = 0. \tag{105}$$

For $v \in N$, we have from (100) and (103) that

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

$$= \begin{cases} 0, & i \in \alpha \\ \max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle, & i \in \beta \\ x_i^*\langle (f_i')'(x^*; d), v\rangle, & i \in \gamma \end{cases} \tag{106}$$

Since N_* is defined in (101) to have the form $\{\xi \in \mathbf{R}^n \,|\, A\xi = 0\}$ for some matrix A, we have that $P_{N_*} w = 0$ if and only if $w = A^T z$ for some z. That is,

$$\frac{1}{2}(\Psi')'(x^*; d)v = \begin{bmatrix} \operatorname{diag} f_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ (\operatorname{diag} x^*_\gamma) f'_{\gamma,\alpha} & (\operatorname{diag} x^*_\gamma) f'_{\gamma,\beta} & (\operatorname{diag} x^*_\gamma) f'_{\gamma,\gamma} \end{bmatrix} \begin{bmatrix} z_\alpha \\ z_\beta \\ z_\gamma \end{bmatrix}, \tag{107}$$

for some $z \in \mathbf{R}^n$. Hence, by matching components from this expression and (106), we have that $P_{N_*}(\Psi')'(x^*; d)v = 0$ if for some $z \in \mathbf{R}^n$ we have

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

Since $z_\alpha = 0$ from the first equation above, and using the definition of the orthonormal matrix Z, we can write these conditions equivalently as follows:

$$[\max(\langle f_i', d\rangle, -d_i)v_i - \min(\langle f_i', d\rangle, -d_i)\langle f_i', v\rangle]_{i\in\beta} = 0, \tag{108a}$$

$$(f_\gamma')'(x^*; d)v = Zt, \quad \text{for some } t \in \mathbf{R}^r. \tag{108b}$$

By the definitions of Z and W, we can restate (108b) equivalently as

$$W^T (f_\gamma')'(x^*; d)v = 0. \tag{109}$$

Since $v \in N$, we have from (100) that

$$v_\alpha = 0, \tag{110a}$$

$$f'_{\gamma,\alpha} v_\alpha + f'_{\gamma,\beta} v_\beta + f'_{\gamma,\gamma} v_\gamma = 0. \tag{110b}$$

The second condition (110b) is equivalent to

$$\begin{bmatrix} Z^T \\ W^T \end{bmatrix} \left[f'_{\gamma,\alpha} \; f'_{\gamma,\beta} \; f'_{\gamma,\gamma} \right] v = 0, \tag{111}$$

Because

$$W^T \left[f'_{\gamma,\alpha} \; f'_{\gamma,\beta} \; f'_{\gamma,\gamma} \right] v = \left[W^T f'_{\gamma,\alpha} \; 0 \; 0 \right] v = W^T f'_{\gamma,\alpha} v_\alpha$$

and $v_\alpha = 0$, the second block row in the system (111) does not add any information and can be dropped. Hence, we can write (110) equivalently as

$$v_\alpha = 0, \qquad Z^T \left[f'_{\gamma,\alpha} \; f'_{\gamma,\beta} \; f'_{\gamma,\gamma} \right] v = 0. \tag{112}$$

By gathering the conditions equivalent to $v \in N$ and $P_{N_*}(\Psi')'(x^*; d)v = 0$, namely, (108a), (109), and (112), we have

$$\begin{bmatrix} \left[e_i^T \right]_{i\in\alpha} \\ [\max(\langle f_i'(x^*), d\rangle, -d_i)e_i - \min(\langle f_i'(x^*), d\rangle, -d_i) f_i'(x^*)]^T_{i\in\beta} \\ Z^T \left[f'_{\gamma,\alpha}(x^*) \; f'_{\gamma,\beta}(x^*) \; f'_{\gamma,\gamma}(x^*) \right] \\ W^T (f_\gamma')'(x^*; d) \end{bmatrix} v = 0,$$

from which we deduce that $v = 0$ whenever the coefficient matrix in this expression is nonsingular. Since the coefficient matrix is precisely the matrix in Definition 9, we have established the result.

Nondegenerate NCP. Considering now the special case of nondegenerate NCP, we obtain a simpler regularity condition, related to 2^{ae}-regularity for nonlinear equations, that ensures that 2^C-regularity holds for the NCP, and hence that the conditions of Theorem 3 are satisfied.

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

$$P_{N^f_{*\gamma}} f'_{\gamma,\gamma}(x^*; d)|_{N^f_\gamma} \tag{113}$$

is nonsingular, where

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

Proof Let the orthonormal matrices W be as in Definition 9, and define two additional orthonormal matrices $\bar{Z}$ and $\bar{W}$ such that the columns of $\bar{W}$ span the null space of $f'_{\gamma,\gamma}$ (and hence the space N^f_γ) and the columns of $\bar{Z}$ span the range space of $(f'_{\gamma,\gamma})^T$. We have that $\bar{Z} \in \mathbf{R}^{|\gamma| \times r}$ and that $\bar{W} \in \mathbf{R}^{|\gamma| \times (|\gamma| - r)}$ and, by the fundamental theorem of linear algebra, that $[\bar{Z} \mid \bar{W}]$ is orthogonal. Hence, 2^C-regularity is equivalent to nonsingularity of the following matrix for almost every $d \in N \setminus \{0\}$:

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

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

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

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 W defined earlier span the subspace $N^f_{*\gamma}$, and since the columns of $\bar{W}$ span the subspace N^f_γ, nonsingularity of $W^T (f'_\gamma)'(x^*; d)\bar{W}$ is equivalent to condition (113).

Nonlinear Equations. We now consider the case in which $\alpha = \beta = \emptyset$, so that the NCP reduced 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 (100) 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^C-regularity is satisfied if

$$P_{N_*^f} f'(x^*; d)|_{N^f} \qquad \text{is nonsingular for almost every } d \in N \setminus \{0\}.$$

This is the 2^{ae}-regularity condition for nonlinear equations (5) introduced by Griewank [8].

NCP with a Modified Weak Regularity Condition. We now consider another special case in which we may have $\beta \neq \emptyset$ and the matrix $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank. The latter 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 in certain important cases.)

Theorem 5 *Suppose that $f'_{\gamma,\beta\cup\gamma}(x^*)$ has full rank $|\gamma|$. Then if the following set of n vectors in $\mathbb{R}^n$ is linearly independent for almost every $d \in N \setminus \{0\}$:*

$$\{e_i\}_{i\in\alpha} \cup \{f'_i(x^*)\}_{i\in\gamma} \cup \{\langle f'_i(x^*), d\rangle e_i + d_i f'_i(x^*)\}_{i\in\beta_1} \cup \{\langle f'_i(x^*), d\rangle f'_i(x^*) + d_i e_i\}_{i\in\beta_2}, \tag{114}$$

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

$$\beta_1(d) = \{i \in \beta \mid \langle f'_i(x^*), d\rangle > -d_i\}, \tag{115a}$$

$$\beta_2(d) = \{i \in \beta \mid \langle f'_i(x^*), d\rangle \le -d_i\}, \tag{115b}$$

then 2^C-regularity is satisfied by the NCP at x^*.

Proof Because of the full-rank assumption, we have $r = |\gamma|$. Using the definitions of the matrices Z and W above, we can set $Z = I$ and W null, so the matrix in Definition 9 reduces to

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

By partitioning the index set β according to (115), we see that the rows of this matrix are identical (to within some changes of sign) to the vectors (114), so the conclusion follows immediately.

The quasi-regularity condition of Izmailov and Solodov [12, Definition 4.1] is somewhat similar to the conditions of Theorem 5; it requires the set (114) to be linearly independent for some $d \in \mathbb{R}^n$ (not restricted to N), for each possible partition $\beta = \beta_1 \cup \beta_2$, with β_1 and β_2 independent of d. (They use a random $d \in \mathbb{R}^n$ in their algorithm.) We now prove a result similar to [12, Proposition 4.2] that gives a sufficient condition for the assumptions of Theorem 5 to hold. This result relies on the following definition of the set $\mathcal{B}$, which is defined to be all partitions of β of the form (115) that are encountered over all possible vectors $d \in N \setminus \{0\}$:

$$\mathcal{B} = \{(\beta_1, \beta_2) \mid \beta_1 = \beta_1(d) \text{ and } \beta_2 = \beta_2(d) \text{ for some } d \in N \setminus \{0\}\}. \tag{116}$$

We make the following assumption, for use in the theorem below.

Assumption 2 *Suppose that for each* $(\beta_1, \beta_2) \in \mathcal{B}$*, there is a vector* $\hat{d} = \hat{d}(\beta_1, \beta_2) \in N$ *such that the set (114) is linearly independent for* $d = \hat{d}$*.*

Theorem 6 *Under Assumption 2 there is a vector* $\tilde{d} \in N \setminus \{0\}$ *(independent of* (β_1, β_2)*) such that (114) is linearly independent with* $d = \tilde{d}$ *for* all $(\beta_1, \beta_2) \in \mathcal{B}$*, and the set of such vectors* $\tilde{d}$ *is open and dense in* N*. In particular, (114) is linearly independent for almost every* $d \in N \setminus \{0\}$*, for all* $(\beta_1, \beta_2) \in \mathcal{B}$*.*

Proof We define U to be a matrix of dimension $n \times |\beta|$ whose columns span the null space of

$$\begin{bmatrix} \left[e_i^T\right]_{i\in\alpha} \\ \left[f_i'(x^*)^T\right]_{i\in\gamma} \end{bmatrix}.$$

(Note that U has $|\beta|$ columns because $\{e_i\}_{i\in\alpha} \cup \{f_i'(x^*)\}_{i\in\gamma}$ is linearly independent, by the assumption.) We then have

$$N = \left\{ Ug \mid g \in \mathbb{R}^{|\beta|} \right\}.$$

With this parametrization of N, the set (114) can be written as follows:

$$\{e_i\}_{i\in\alpha} \cup \{f_i'(x^*)\}_{i\in\gamma} \cup \{\langle f_i'(x^*), Ug\rangle e_i + (Ug)_i f_i'(x^*)\}_{i\in\beta_1} \cup \\ \{\langle f_i'(x^*), Ug\rangle f_i'(x^*) + (Ug)_i e_i\}_{i\in\beta_2}.$$

Note that linear independence of this set is equivalent to nonzeroness of the determinant of the $n \times n$ matrix formed by these vectors, and that the determinant is a polynomial function of the components of g.

Given $\hat{d} = U\hat{g}$ for which the set (114) is linearly independent, we note as in [12] that the determinant is nonzero for $g = \hat{g}$, thus is nonzero for all $g \in \mathbb{R}^{|\beta|}$ except for a set of measure zero. A similar observation can be made for each $(\beta_1, \beta_2) \in \mathcal{B}$. By taking the intersection of these finitely many subsets of $\mathbb{R}^{|\beta|}$, all of which exclude only a measure-zero set of points in $\mathbb{R}^{|\beta|}$, we obtain a set containing all of $\mathbb{R}^{|\beta|}$ except for a set of measure zero. By applying U, the set in question transforms to the whole set N except for a set of measure zero, as claimed.

The following result is similar to [12, Proposition 4.3].

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

The proof of the first claim is nearly identical to the proof of Proposition 4.3 [12]. (In the notation of [12], restrict p to N and note that $L \equiv N_\perp$.) Full rank of $f'_{\gamma,\beta\cup\gamma}(x^*)$ follows directly from b-regularity, while 2^C-regularity follows from Theorems 6 and 5. We prove a partial converse of this result by modifying and extending the argument given by Izmailov and Solodov [12, p. 400].

Proposition 4 *Suppose that x^* is not a b-regular solution of the NCP, and that $|\beta| = 1$. Then it cannot be true that the set (114) is linearly independent for almost every $d \in N \setminus \{0\}$.*

Proof Consider the set

$$\{e_i\}_{i\in\alpha} \cup \{f_i'(x^*)\}_{i\in\gamma}, \tag{117}$$

and denote the range of this set by L. Let $\beta = \{i_0\}$, and assume that b-regularity does not hold. If (117) is rank deficient then (114) is not linearly independent 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 $L \equiv N_\perp$ we have for all $d \in N$ that $\langle f_{i_0}'(x^*), d\rangle = 0$. Therefore from (115), we have

$$d_{i_0} > 0 \;\Leftrightarrow\; \beta_1(d) = \{i_0\}, \quad \beta_2(d) = \emptyset \tag{118a}$$

$$d_{i_0} \le 0 \;\Leftrightarrow\; \beta_1(d) = \emptyset, \quad \beta_2(d) = \{i_0\}. \tag{118b}$$

In the case (118a), the set (114) is linearly dependent 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 (114) to be linearly independent for almost all $d \in N$, we must have (118b) 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

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

for almost all $d \in N$, so that (114) is linearly dependent 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 (115) we have

$$\langle f_{i_0}'(x^*), d\rangle > 0 \;\Leftrightarrow\; \beta_1(d) = \{i_0\}, \quad \beta_2(d) = \emptyset \tag{119a}$$

$$\langle f_{i_0}'(x^*), d\rangle \le 0 \;\Leftrightarrow\; \beta_1(d) = \emptyset, \quad \beta_2(d) = \{i_0\}. \tag{119b}$$

In case (119a), the set (114) is linearly dependent 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 (114) to be linearly independent, we must be in the case (119b) 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

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

for almost all $d \in N$, so that (114) is linearly dependent 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^C-regularity and b-regularity are equivalent. However, when $f_{\gamma,\beta\cup\gamma}'(x^*)$ is rank deficient, 2^C-regularity may still hold, even though b-regularity fails.

In the case of nondegenerate NCP—that is, $\beta = \emptyset$ and $f_{\gamma,\gamma}'(x^*)$ nonsingular—we have from (100) and (101) that $N = N_* = \{0\}$, so that x^* is a nonsingular solution of $\Psi(x^*) = 0$ and Theorem 3 does not apply.

Table 1 Convergence rate of Newton's Method on Ψ_S for the Simple NCP test problems, showing regularity properties. (• = 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	—	—	•	—	—
aff1	2	0	suplin	1	0	1	•	•	•	—	—
DIS61, 2	2	0	suplin	1	0	1	•	•	•	—	—
quarquad, 1	2	1	1/2	0	1	1	•	•	•	•	•
affknot1	2	1	1/2	0	1	1		•			
affknot2	2	1	1/2	0	1	1	•	•	•	•	•
quadknot	2	2	1/2	0	1	1					•
munson4	2	2	1/2	0	0	2				•	•
DIS61, 1	2	2	1/2	0	1	1				•	•
DIS64	2	2	1/2	0	2	0	—	—	•	•	•
nehard	3	2	1/2	0	2	1	•	•			•
quad1	2	1	2/3	0	1	1	•	•			
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					

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. 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.

We also the numbers of iterations required for local convergence of Newton's method and the Accelerated-Newton method of Section 5 using the subset of Simple NCP test problems with 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 is the only problem in this subset with a convergence rate of $\frac{1}{2}$ for Newton's method but without 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 $\|\Psi(x)\|_2 \leq 10^{-11}$.

Table 2 shows the number of iterations required by the Newton and Accelerated Newton methods for the subset of problems discussed above. The final column shows the number of steps taken in the "accelerated phase," following detection of a linear convergence rate in the pure Newton method.

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
nehard	25	19	5

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 mean convergence rate in the accelerated phase as predicted by Theorem 2 is $\sqrt{\frac{1}{2}(1-\frac{\alpha}{2})} \approx 0.158$, which 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 it is worth stating in full here.

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

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

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 $\bar{\mathcal{R}}$ to the point $\bar{x}$.

$$\bar{x} = x - F'(x)^{-1}F(x). \tag{120}$$

We drop the dependence on x in the following.

By the definition of $\bar{\mathcal{R}}$, we have that $\sigma(t) > 0$. In the remainder of this discussion, we drop the argument t from $\sigma(t)$ for clarity. Using positivity of σ and (12), 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

$$G^{-1}(x) = \rho^{-1}\bar{B}^{-1}(t) + \sigma^{-2}O(\rho^0) = \sigma^{-2}O(\rho^{-1}), \tag{121}$$

(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 (12) 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 (12), (121), and the positivity of σ, we have

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

Hence,

$$(122) \qquad F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I + \|G^{-1}\|O(\rho^2) & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) + \sigma^{-2}O(\rho) + \|G^{-1}\|O(\rho^2) \\ O(\rho) + \|G^{-1}\|O(\rho^3), & I + O(\rho) + \|G^{-1}\|\rho^2 \end{bmatrix} x.$$

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

$$(123) \qquad F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & -\frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & I \end{bmatrix} x - e(x),$$

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

$$(124) \qquad \|e(x)\| \le \delta \frac{\|x\|^2}{\sigma(x/\|x\|)^2} = \delta \frac{\rho^2}{\sigma^2},$$

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 (12) which, as we have already noted, are finite.) The definition of $r(t)$ (30) uses this value of δ.

Using (123), we have

$$(125) \qquad \bar{x} = x - F'(x)^{-1}F(x) = \begin{bmatrix} \frac{1}{2}I & \frac{1}{2}\bar{B}^{-1}(t)\bar{C}(t) \\ 0 & 0 \end{bmatrix} x + e(x) = \frac{1}{2}g(x) + e(x),$$

where $g(x)$ is defined in (28).

Here we follow Griewank in implicitly assuming that $\sigma \le 1$. (If necessary, we can replace σ with $\min(\sigma, 1)$ in the analysis to ensure that this property holds.)

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

Denoting the sequence of Newton iterates by $\{x_j\}_{j\ge0}$, we use the following associated notation in the remainder of this section:

$$(126) \qquad \rho_j = \|x_j\|, \quad t_j = x_j/\rho_j, \quad \sigma_j = \sigma(t_j), \quad s_j = g(x_j)/\|g(x_j)\|.$$

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

$$(127) \qquad \psi_j(s) = \cos^{-1} t_j^T s.$$

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 $v \in \mathsf{R}^n$ and $s \in \mathcal{S}$. This angle must lie in the range $[0, \pi]$. If $\omega_{v,s} \le \pi/2$, we have

$$\sin \omega_{v,s} = \min_{\lambda \in \mathsf{R}} \|\lambda v - s\|, \tag{128}$$

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

By applying (125) to the Newton step from x_j to x_{j+1}, we have

$$\left\| x_{j+1} - \frac{1}{2} g(x_j) \right\| = \left\| x_j - F'(x_j)^{-1} F(x_j) - \frac{1}{2} g(x_j) \right\| = \|e(x_j)\| \le \delta \frac{\rho_j^2}{\sigma_j^2}, \tag{129}$$

By setting $j = 0$ in (129) and using $x_0 \in \mathcal{R}$, we have

$$\sin \psi_1(s_0) = \min_{\lambda \in \mathsf{R}} \left\| \lambda x_1 - \frac{g(x_0)}{\|g(x_0)\|} \right\| \le \left(\frac{1}{2} \|g(t_0)\| \right)^{-1} \delta \frac{\rho_0}{{\sigma_0}^2}. \tag{130}$$

The equality in (130) is a consequence of (128), provided that $\psi_1(s_0) \le \pi/2$. (We verify the latter fact in Appendix B.) By the third part of the definition of $r(t)$ (30), we have from (130) 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 (129) by ρ_{j+1}, we obtain

$$\sin \theta_{j+1} = \min_{y \in N} \|t_{j+1} - y\| \le \delta \frac{\rho_j^2}{\sigma_j^2 \rho_{j+1}}, \quad \text{for } j = 0, 1, 2, \dots . \tag{131}$$

The equality in (131) is valid because N is a linear subspace. (For the starting point we have only the trivial upper bound $\sin \theta_0 \le 1$.) By the definition of g (28), we have

$$\begin{aligned} x_j - g(x_j) &= \begin{bmatrix} x_j \cos \theta_j \\ x_j \sin \theta_j \end{bmatrix} - \begin{bmatrix} I & \bar{B}^{-1}(t_j)\bar{C}(t_j) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_j \cos \theta_j \\ x_j \sin \theta_j \end{bmatrix} \\ &= \begin{bmatrix} -\bar{B}^{-1}(t_j)\bar{C}(t_j) x_j \sin \theta_j \\ x_j \sin \theta_j \end{bmatrix}. \end{aligned} \tag{132}$$

By combining (129) and (132), we obtain

$$\begin{aligned} \left\| x_{j+1} - \frac{1}{2} x_j \right\| &\le \left\| \frac{1}{2}(x_j - g(x_j)) \right\| + \left\| x_{j+1} - \frac{1}{2} g(x_j) \right\| \\ &\le \frac{1}{2} \left(\| - \bar{B}^{-1}(t_j)\bar{C}(t_j)\| + 1 \right) \|x_j \sin \theta_j\| + \delta \frac{{\rho_j}^2}{\sigma_j^2} \\ &\le \left(\frac{1}{2} \frac{\|\bar{C}(t_j)\| + \sigma_j}{\sigma_j} \sin \theta_j + \delta \frac{\rho_j}{\sigma_j^2} \right) \rho_j \\ &\le \left(\frac{1}{2} \frac{c}{\sigma_j} \sin \theta_j + \delta \frac{\rho_j}{\sigma_j^2} \right) \rho_j . \end{aligned} \tag{133}$$

Dividing by ρ_j and applying the inverse triangle inequality, we find that

$$\left| \frac{\rho_{j+1}}{\rho_j} - \frac{1}{2} \right| \le \frac{1}{2} \frac{c}{\sigma_j} \sin \theta_j + \delta \frac{\rho_j}{{\sigma_j}^2}. \tag{134}$$

Upon setting j to zero and rearranging (134), we obtain

$$\rho_1 \leq \rho_0 \left(\frac{1}{2} \left(1 + \frac{c}{\sigma_0} \right) + \delta \frac{\rho_0}{{\sigma_0}^2} \right). \tag{135}$$

By the second part of the definition (30), we have

$$\rho_0 \leq \frac{\sigma_0 \hat{\rho}(s_0)}{c + \sigma_0},$$

and also

$$\rho_0 \leq \frac{\sigma_0^2 \hat{\rho}(s_0)}{2\delta r_b} \leq \frac{\sigma_0^2 \hat{\rho}(s_0)}{2\delta \rho_0}.$$

Applying these inequalities to (135) yields

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

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

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

Griewank gives an inductive argument for linear convergence in the proximity of N in Section 5 of [8]. We give the argument here for future reference.

As in (126), 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 $\{x_j = \rho_j t_j\}_{j \geq 1}$ maintains the properties

$$\rho_j < \hat{\rho}(s_0) \equiv \hat{\rho}, \quad \theta_j \leq \hat{\phi}(s_0) \equiv \hat{\phi}, \quad \psi_j(s_0) \equiv \psi_j < \phi(s_0) \equiv \phi. \tag{136}$$

By the first and third properties, the iterates remain in $\bar{\mathcal{R}}$. Further, because of (21), the third property implies that

$$\sigma_j \equiv \sigma(t_j) \geq \hat{\sigma}(s_0) \equiv \hat{\sigma} > 0, \tag{137}$$

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

$$q \equiv q(s_0). \tag{138}$$

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

We assume that (136) holds for all $1 \leq i \leq j$. We have

$$\begin{aligned}
\left| \frac{\rho_{j+1}}{\rho_j} - \frac{1}{2} \right| &\leq \frac{1}{2}\frac{c}{\sigma_j} \sin\hat{\phi} + \delta \frac{\hat{\rho}}{{\sigma_j}^2} && \text{from (134) and (136)} \\
&\leq \left(\frac{1}{2}\frac{c}{\sigma_j} + \frac{(1-q)\hat{\sigma}^2}{2{\sigma_j}^2} \right) \sin\hat{\phi} && \text{from (25)} \\
&\leq \frac{c/\sigma_j + 1 - q}{2} \sin\hat{\phi} && \text{from (135)} \\
&\leq \frac{q}{2}, && \text{from (24).}
\end{aligned}$$

Equivalently,

$$\frac{1-q}{2} \leq \frac{\rho_{i+1}}{\rho_i} \leq \frac{1+q}{2}, \quad \text{for } i = 1, 2, \ldots, j. \tag{139}$$

From the right inequality of (139), we have

$$\rho_{i+1} \le \rho_1 \left(\frac{1+q}{2}\right)^i < \hat{\rho}, \quad \text{for } i = 1, 2, \ldots, j. \tag{140}$$

From (131), we have

(141)

$$\begin{aligned}
\sin\theta_{i+1} &\le \frac{\delta\rho_i^2}{\sigma_i^2\rho_{i+1}} && \text{for } i = 1, 2, \ldots, j \\
&\le \left(\frac{\delta\rho_i}{\sigma_i^2}\right)\left(\frac{2}{1-q}\right) && \text{for } i = 1, 2, \ldots, j, \text{ by the left inequality in (139)} \\
&< \left(\frac{\delta\hat{\rho}}{\hat{\sigma}^2}\right)\left(\frac{2}{1-q}\right) && \text{for } i = 1, 2, \ldots, j, \text{ by (136)} \\
&= \sin\hat{\phi} && \text{for } i = 1, 2, \ldots, j, \text{ by the definition of } \hat{\rho} \text{ (25).}
\end{aligned}$$

Therefore $\theta_{j+1} < \hat{\phi}$.

Let $\Delta\psi_j$ be the angle between two consecutive iterates x_j and x_{j+1}. From (133) we have the upper bound

$$\sin\Delta\psi_i = \min_{\lambda\in\mathbf{R}} \|\lambda x_{i+1} - t_i\| \le \frac{c}{\sigma_i}\sin\theta_i + \frac{2\delta\rho_i}{\sigma_i^2}, \quad \text{for } i = 1, 2, \ldots, j. \tag{142}$$

In Appendix B, we verify the equality in (142) 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

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

From (142), we have

$$\sum_{i=1}^{j} \sin\Delta\psi_i \le \frac{c}{\hat{\sigma}}\left(\sin\theta_1 + \sum_{i=1}^{j-1}\sin\theta_{i+1}\right) + \frac{2\delta}{\hat{\sigma}^2}\sum_{i=1}^{j}\rho_i. \tag{143}$$

Using (140), the bound $\rho_1 \le \hat{\rho}$, and the definition of $\hat{\rho}$ (25), 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 (141), 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 (143), we have

$$\sum_{i=1}^{j} \sin \Delta\psi_i \leq \frac{c}{\hat{\sigma}} \left(\sin\theta_1 + \frac{2}{(1-q)} \sin\hat{\phi} \right) + 2\sin\hat{\phi}. \tag{144}$$

By adding $\sin\hat{\phi}$ to this sum and using $\theta_1 \leq \hat{\phi}$ and the first inequality implicit in the definition of $\sin\hat{\phi}$ (24), we have

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

Since $q = \frac{1}{4}\sin\phi \leq \frac{1}{4}$, we find that $\sin\psi_{j+1} < \sin\phi$ and thus $\psi_{j+1} < \phi$.

This observation shows that the 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 (27).

Noting that (140) and (141) 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 (134), 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 (128)

In this section, we justify the use of the formula (128) 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 $\alpha_{v/s}s$ 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 (130). 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 (129), 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 $r(t)$ (30), 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 (142) by showing that $\alpha_{x_{i+1}/t_i} > 0$. By (133), we have for $i \in \{1, 2, \ldots, j\}$,

$$\begin{aligned} \left\| x_{i+1} - \frac{1}{2} x_i \right\| &\leq \left(\frac{1}{2} \frac{c}{\sigma_i} \sin \theta_i + \delta \frac{\rho_i}{\sigma_i^2} \right) \rho_i \\ &< \left(\frac{1}{2} \frac{c}{\hat{\sigma}} \sin \hat{\phi} + \delta \frac{\hat{\rho}}{\hat{\sigma}^2} \right) \rho_i && \text{by (136)} \\ &= \frac{1}{2} \left(\frac{c}{\hat{\sigma}} + 1 - q \right) \sin \hat{\phi} \rho_i && \text{by (25)} \\ &\leq \frac{q}{2} \rho_i && \text{by (24).} \end{aligned}$$

Hence, 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 (23).

Third, we verify (63) by showing that $\Delta\psi_1 < \pi/2$. From (64),

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

Hence

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

which implies

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

where we have used the definition of q_α (36) 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 (83) to justify (90) as follows. From (83), 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 (94). Since (75) 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 $\alpha_{x_{2j+1}/t_{2j}} t_{2j} \geq \frac{1}{2}(1 - q_\alpha)\rho_{2j} > 0$ 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^* if it unique or ${x^*}^1$, ${x^*}^2$, etc. if there are multiple solutions. The starting point used for solution ${x^*}^i$ is denoted below as ${x_0}^i$.

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

Table 3 Starting Point/Local 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)
nehard	(10, 1, 10)	$(0, 0, \sqrt{200})$
quad1	(0.9, 0.1)	(1, 0)
quarquad, 2	(0.9, 0.1)	(1, 0)
quarp, 2	0.9	1
quarn	0.9	1

8. munson4 (from MCPLIB [16])

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

9. DIS64 ([1, Example 6.4])

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

10. nehard (from MCPLIB [16])

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

11. quad1

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

12. quarn

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

Acknowledgements

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. Kanzow, C.: Some equation-based methods for the nonlinear complementarity problem. Optimization Methods and Software **3**(1), 327–340 (1994)
15. 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)
16. ftp://ftp.cs.wisc.edu/math-prog/mcplib
17. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Mathematical Programming **58**, 353–367 (1993)
18. Reddien, G.W.: On Newton's method for singular problems. SIAM Journal on Numerical Analysis **15**(5), 993–996 (1978)
19. Reddien, G.W.: Newton's method and high order singularities. Computers & Mathematics with Applications **5**, 79–86 (1979)