Research Article

On the Solution of a Class of Nonlinear Systems Governed by an $M$-Matrix

Woula Themistoclakis and Antonia Vecchio

Istituto per le Applicazioni del Calcolo “Mauro Picone”, Sede di Napoli, National Research Council of Italy (CNR), Via P. Castellino, 111, 80131 Napoli, Italy

Correspondence should be addressed to Antonia Vecchio, antonia.vecchio@cnr.it

Received 15 November 2011; Accepted 13 March 2012

Copyright © 2012 W. Themistoclakis and A. Vecchio. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider a weakly nonlinear system of the form $(I + \varphi(x)A)x = p$, where $\varphi(x)$ is a real function of the unknown vector $x$, and $(I + \varphi(x)A)$ is an $M$-matrix. We propose to solve it by means of a sequence of linear systems defined by the iteration procedure $(I + \varphi(x)A)x_{r+1} = p, r = 0, 1, \ldots$. The global convergence is proved by considering a related fixed-point problem.

1. Introduction

In recent years [1, 2], we approached the numerical resolution of the following nonstandard integrodifferential problems arising in the study of the kinetic theory of dusty plasmas:

$$
\nu(t)y(t) = \int_0^\infty k(s)y(s)ds\left(D(t)y'(t)\right)' + p(t), \quad t > 0, \quad y'(0) = 0, \quad y(\infty) = 0,
$$

(1.1)

whose discretization, by means of a difference scheme and a quadrature rule [2], leads to a particular kind of nonlinear system of equations. Solving it by means of a fixed-point (FP) iteration process, we noted that such a procedure seems to globally converge, that is, it converges independently of the choice of the starting point. Our aim, here, is to explain the reason of this “nice” behavior. Throughout the paper, the notation $M \geq 0$, with $M \in \mathbb{R}^{N \times K}$, means that each element of $M$ is nonnegative, whereas $M > 0$ means $M \geq 0$ and $M \neq 0$. The nonlinear system under investigation is

$$
(I + \varphi(x)A)x = p,
$$

(1.2)
where \( I \) is the \( N \)-dimensional identity matrix, \( x \in \mathbb{R}^N \) is an unknown vector, and we make the following assumptions which will be sufficient for the global convergence of the iterative process we propose (see Theorem 3.3):

(i) \( p \in \mathbb{R}^N \) with \( p > 0 \);

(ii) \( A \in \mathbb{R}^{N \times N} \) has nonnegative main diagonal and nonpositive off-diagonal entries (Z-matrix);

(iii) \( A \) is rowwise weakly diagonal dominant;

(iv) \( \varphi : \mathbb{R}^N \to \mathbb{R} \) is a differentiable function such that \( \varphi(x) \geq 0 \), for all \( x \geq 0 \) and it is homogeneous of degree \( \gamma \leq 1 \), that is, for all \( \alpha > 0 \), \( \varphi(\alpha x) = \alpha^\gamma \varphi(x) \);

(v) \( \nabla \varphi(x) > 0 \) for all \( x \geq 0 \), where \( \nabla \varphi(x) \) denotes the (row vector) gradient of \( \varphi(x) \).

Systems of type (1.2) fall into the class of weakly nonlinear systems (see, e.g., [3, 4]) and they also arise from the discretization, by finite-difference methods, of mathematical models less cumbersome than (1.1), as, for instance, the differential problems of the form:

\[
y(t) + b(y)a(t)y'(t) = p(t), \quad t \geq 0, \quad y(0) = 0.
\]

(1.3)

A central rule in our results is played by the matrix \( (I + \varphi(x)A) \) appearing in (1.2) which, as it can be immediately seen from the assumption (ii)-(iv), is an \( M \)-matrix for all \( x \geq 0 \) (see, e.g., [5, page 137, prop. \( M_{35} \)]).

The organization of the paper is the following. In Section 2, we investigate a particular class of one variable functions \( f \) such that the FP iterations \( t_{r+1} = f(t_r) \) converge independently of the choice of the starting value \( t_0 \). By exploiting this result jointly with the properties of the matrix \( I + \varphi(x)A \), we are led to state that, under the assumptions (i)–(v), the FP iteration procedure \((I + \varphi(x_r)A)x_{r+1} = p\) globally converges to a solution of (1.2). This constitutes our main result which is contained in Section 3. Finally, some numerical experiments showing the sharpness of the required conditions and the performance of the iteration scheme are reported in Section 4.

### 2. The One Variable Case

In order to prove our main theorem, we need the following result on a particular class of numerical sequences.

**Theorem 2.1.** Let \( f \) be a one variable scalar function such that

(a) \( f : t \in X \subseteq [0, +\infty) \to (0, M_f) \), with \( M_f > 0 \), \( X \) closed and \( f(X) \subset X \);

(b) \( f \) continuous;

(c) \( tf(t) \) strictly increasing.

Then, starting from any \( t_0 \in X \), the sequence obtained by the functional iteration process

\[
t_{r+1} = f(t_r), \quad r = 0, 1, 2, \ldots,
\]

(2.1)

converges to \( t^* \), a fixed point of \( f \).

In this section, we are going to prove this theorem.
Denote by $g$ the function $g(t) = t - f(t)$, then we have
\[ g(t_r) = t_r - f(t_r) = t_r - t_{r+1}. \] (2.2)

Of course, if there exists a point $t_r$ such that $g(t_r) = 0$, then $t_r = t_{r+n}$ for all $n = 1, \ldots$ and the statement is trivial. Similarly, if the sequence $g(t_r)$ is ultimately nonnegative (nonpositive), we have that the sequence $t_r$ is ultimately decreasing (increasing) and it trivially converges by hypothesis (a). Therefore, from now on, we assume that the sequence $g(t_r)$ satisfies
\[ \forall r \geq 0, \exists m > r : g(t_r)g(t_m) < 0. \] (2.3)

The following lemma holds.

**Lemma 2.2.** If $g(t_r) < 0$ (resp. $> 0$) and $g(t_{r+m}) > 0$ (resp. $< 0$), $m = 1, \ldots, n$, then
\[ t_r < t_{r+m+2} \ (\text{resp. } t_r > t_{r+m+2}), \quad m = -1, 0, 1, \ldots, n. \] (2.4)

**Proof.** Let us prove (2.4) by induction on $n \geq 1$. Assume $n = 1$, that is, we are assuming that
\[ g(t_r) < 0, \quad g(t_{r+1}) > 0, \] (2.5)
and we want to prove
\[ t_r < t_{r+m+2}, \quad m = -1, 0, 1. \] (2.6)
From the first of (2.5) and (2.2), we immediately get (2.6) with $m = -1$, that is, we have
\[ t_r < t_{r+1}, \] (2.7)
and so, from the hypothesis (c), we deduce
\[ t_r f(t_r) < t_{r+1} f(t_{r+1}), \] (2.8)
which, recalling that $f(t_r) = t_{r+1}$, implies
\[ t_r t_{r+1} < t_{r+1} t_{r+2}, \] (2.9)
and, therefore, (2.6) holds with $m = 0$, that is,
\[ t_r < t_{r+2}. \] (2.10)
By using (c) once again, (2.10) gives $t_r f(t_r) < t_{r+2} f(t_{r+2})$ or equivalently
\[ t_r t_{r+1} < t_{r+2} t_{r+3}. \] (2.11)
But the second of (2.5) implies \( t_{r+2} < t_{r+1} \), so that from (2.11), we obtain
\[
t_{r}t_{r+2} < t_{r}t_{r+1} < t_{r+2}t_{r+3},
\]
which assures that (2.6) holds for \( m = 1 \) too.

Now, we assume that the lemma holds for \( m = n \), that is, suppose that
\[
g(t_{r}) < 0, \quad g(t_{r+m}) > 0, \quad m = 1, \ldots, n,
\]
implies
\[
t_{r} < t_{r+m+2}, \quad m = -1, 0, 1, \ldots, n.
\]

In order to prove the lemma for \( m = n + 1 \), assume
\[
g(t_{r}) < 0, \quad g(t_{r+m}) > 0, \quad m = 1, \ldots, n + 1.
\]

Of course (2.14) holds, and the second of (2.15) implies
\[
t_{r+1} > t_{r+2} > \cdots > t_{r+n+2}.
\]

But from (2.14) with \( m = n \) and (c), we have
\[
t_{r}t_{r+1} < t_{r+n+2}t_{r+n+3},
\]
which in view of (2.16) leads to
\[
t_{r}t_{r+n+2} < t_{r+n+2}t_{r+n+3}.
\]

Hence, \( t_{r} < t_{r+n+3} \), that is, (2.14) holds for \( m = n + 1 \) too, and the desired result is proved. \( \square \)

Now, let us partition the sequence \( \{t_{r}\} \) into two subsequences \( \{t_{ri}\}_{i} \) and \( \{t_{sj}\}_{j} \) defined by
\[
g(t_{ri}) < 0, \quad g(t_{sj}) > 0, \quad i, j = 1, 2, \ldots
\]

In order to state Theorem 2.1, we are going to prove that these subsequences converge to the same limit. First of all, the following result holds.

**Lemma 2.3.** The sequences \( \{t_{ri}\}_{i} \) and \( \{t_{sj}\}_{j} \) are strictly increasing and decreasing respectively.

**Proof.** Let us prove that the sequence \( \{t_{ri}\} \) is strictly increasing, the proof about \( \{t_{sj}\} \) is analogous, and it is omitted. Consider \( r_{i} \) such that \( g(t_{ri}) < 0 \). Then, two cases may occur: (I) \( g(t_{ri+1}) < 0 \) or (II) \( g(t_{ri+1}) > 0 \).
In this case, it is clear that $t_{r+1}$ belongs to the same subsequence $\{t_r\}_i$ and we immediately have

$$t_r < t_{r+1} = t_{r+1}.$$  \hspace{1cm} (2.20)

(II) From (2.3), there exists $n \geq 2$ such that $t_{r+1} = t_{r+n}$, that is, we have

$$g(t_r) < 0, \quad g(t_{r+m}) > 0 \quad \text{for} \quad m = 1, \ldots, n-1, \quad g(t_{r+n}) < 0. \hspace{1cm} (2.21)$$

Then, Lemma 2.2 with $r = r_i$ assures that

$$t_r < t_{r+2r} = m = -1, 0, 1, \ldots, n-1,$$ \hspace{1cm} (2.22)

which, for $m = n-2 \geq 0$, gives

$$t_r < t_{r+n} = t_{r+1}.$$ \hspace{1cm} (2.23)

The desired result comes out from (2.20) and (2.23).

**Lemma 2.4.** The sequences $\{t_r\}_i$ and $\{t_{s_j}\}_j$ provide two separate sets, that is,

$$t_r \leq t_{s_j} \quad \forall i, j \in \mathbb{N}. \hspace{1cm} (2.24)$$

**Proof.** Suppose, for example, $r_i < s_j$ and precisely let $s_j = r_i + k$. If $k = 1$, we get (2.24) directly from the assumption $0 > g(t_r) = t_r - t_{r+1}$. If $k \geq 2$, observe that the numerical sequence $\{g(t_r), g(t_{r+1}), \ldots, g(t_{r+k})\}$ presents at least one change of sign, in view of (2.19).

Let us examine the case of an unique change of sign, by supposing that there exists an integer $l$, with $0 \leq l \leq k-1$, such that

$$g(t_{r+m}) < 0 \quad \text{for} \quad m = 0, 1, \ldots, l, \quad g(t_{r+l+m}) > 0 \quad \text{for} \quad m = 1, \ldots, k-l, \hspace{1cm} (2.25)$$

the other cases being analogous.

By Lemma 2.3, we get

$$t_r \leq t_{r+1} \leq \cdots \leq t_{r+l}, \quad t_{r+k} \leq t_{r+k-1} \leq \cdots \leq t_{r+l+1},$$ \hspace{1cm} (2.26)

and by applying Lemma 2.2 with $r = r_i + l$, we deduce

$$t_{r+l} < t_{r+l+m+2}, \quad m = -1, 0, 1, \ldots, k-l.$$ \hspace{1cm} (2.27)

Hence, by taking $m = k-l-2 \geq -1$, we conclude

$$t_r < t_{r+1} < \cdots < t_{r+l} < t_{r+k} < t_{r+k-1} < \cdots < t_{r+l+1}.$$ \hspace{1cm} (2.28)
By the previous lemmas, we get the next result which completes the proof of the convergence of the FP iteration process.

**Theorem 2.5.** For any choice of \( t_0 \in X \), the sequence \( \{ t_r \} \), defined in (2.1) converges.

**Proof.** It is sufficient to prove that the subsequences \( \{ t_{r_i} \} \) and \( \{ t_{s_j} \} \) defined in (2.19) converge to the same limit, satisfying

\[
\lim i t_{r_i} = \lim j t_{s_j} = \lim r t_r. \tag{2.29}
\]

By the previous lemmas, we partitioned the sequence \( \{ t_r \} \) into two subsequences, \( \{ t_{r_i} \} \) and \( \{ t_{s_j} \} \), which are separate, strictly monotone, and bounded sequences. Hence, they converge and

\[
0 \leq \sup i t_{r_i} = \lim i t_{r_i} = l_1 \leq l_2 = \lim j t_{s_j} = \inf j t_{s_j} \leq M_f. \tag{2.30}
\]

Assume ab absurdo that

\[
l_1 < l_2. \tag{2.31}
\]

Let us consider the subsequence \( \{ t_{r_i} \} \) of \( \{ t_{r_i} \} \) and the subsequence \( \{ t_{s_j} \} \) of \( \{ t_{s_j} \} \), such that

\[
g(t_{r_i-1}) > 0, \quad g(t_{r_i}) < 0, \quad g(t_{s_j-1}) < 0, \quad g(t_{s_j}) > 0. \tag{2.32}
\]

In other words, \( t_{r_i} \) and \( t_{s_j} \) are those elements of \( \{ t_{r_i} \} \) and \( \{ t_{s_j} \} \), respectively, whose previous elements, in the main sequence \( \{ t_r \} \), belong to the other subsequence \( \{ t_{s_j} \} \) and \( \{ t_{r_i} \} \) respectively, that is, we have that

\[
\exists k \in \mathbb{N} \text{ such that } t_{r_i-1} = t_{s_k}, \quad \exists h \in \mathbb{N} \text{ such that } t_{s_j-1} = t_{r_h}. \tag{2.33}
\]

Of course, as \( \{ t_{r_i} \} \) is a subsequence of the convergent sequence \( \{ t_{r_i} \} \), there results

\[
\lim i t_{r_i} = l_1, \tag{2.34}
\]

and the same is true for \( \{ t_{s_j} \} \), that is,

\[
\lim j t_{s_j} = l_2. \tag{2.35}
\]
where \( l_1 \) and \( l_2 \) are given in (2.30). On the other hand, recalling that \( t_{\rho_i} = f(t_{\rho_i-1}) \), from (b) and (2.33), we have

\[
\begin{align*}
l_1 &= \lim_{i} t_{\rho_i} = \lim_{i} f(t_{\rho_i-1}) = \lim_{k} f(t_{\alpha_i}) = f(l_2), \\
l_2 &= \lim_{j} t_{\sigma_j} = \lim_{j} f(t_{\sigma_j-1}) = \lim_{k} f(t_{\rho_k}) = f(l_1).
\end{align*}
\]

(2.36)

But by the hypothesis (c), the assumption (2.31) implies \( l_1 f(l_1) < l_2 f(l_2) \), that is, in view of (2.36), we have \( l_1 l_2 < l_2 l_1 \), which is absurd. Thus, the result \( l_1 = l_2 = \lim_{r} t_r \) is achieved. \( \square \)

### 3. The Solution of the Nonlinear System

In this section, we come back to the nonlinear system (1.2). First of all, we recall that, under the hypotheses (ii)--(iv), for any \( x \geq 0 \), the matrix \((I + \varphi(x)A)\) appearing in (1.2) is an M-matrix and it satisfies (see, e.g., [5, prop. N38])

\[
(I + \varphi(x)A)^{-1} > 0, \quad \forall x \geq 0.
\]

(3.1)

Now, we define the function \( F \) as

\[
F : q \in [0, +\infty) \rightarrow \varphi\left((I + qA)^{-1}p\right),
\]

(3.2)

and we are going to prove that it satisfies the properties stated in the following theorem, where \( \gamma \) is the homogeneity degree of \( \varphi \) appearing in (iv) and the classical notation \( e = (1, \ldots, 1)^T \in \mathbb{R}^N \) is used.

**Theorem 3.1.** Assume that (i)--(v) hold. Then, \( F \) is a continuous, differentiable function, and for any \( q \geq 0 \), it satisfies

\[
F(q) \in (0, M_F) \quad \text{with} \quad M_F = \|p\|_\infty \varphi(e), \quad \text{(3.3)}
\]

\[
F'(q) = -\text{grad} \left( \varphi\left((I + qA)^{-1}p\right)\right)(I + qA)^{-1}A(I + qA)^{-1}p, \quad \text{(3.4)}
\]

\[
s(q) := q^T F(q) \quad \text{strictly increasing.} \quad \text{(3.5)}
\]

**Proof.** As the inverse of \((I + qA)\) is defined for all \( q \geq 0 \), then \( F \) is clearly a continuous function. Moreover, it can be easily seen that

\[
\frac{d}{dq} (I + qA)^{-1} = -(I + qA)^{-1}A(I + qA)^{-1},
\]

(3.6)

hence, (3.4) is immediately true. In view of Theorem A.1 in [6], we observe that (ii) and (iii) assures

\[
\| (I + qA)^{-1} \|_\infty \leq 1, \quad \forall q \geq 0.
\]

(3.7)
Therefore, for any $q \geq 0$, the vector $x(q) := (I + qA)^{-1}p$ satisfies $\|x(q)\|_{\infty} \leq \|p\|_{\infty}$ and because of (v), we get $F(q) = \varphi(x(q)) \leq \varphi(\|p\|_{\infty}e)$. Consequently, assertion (3.3) holds taking into account (iv). Finally, again in view of (iv), we have

$$s(q) := q^tF(q) = q^t\varphi(x(q)) = \varphi(qx(q)) \implies s'(q) = \text{grad} (\varphi(qx(q)))(x(q) + qx'(q)), \tag{3.8}$$

where, recalling (3.6), it results

$$x(q) + qx'(q) = [(I + qA)^{-1} - (I + qA)^{-1}qA(I + qA)^{-1}]p = (I + qA)^{-2}p. \tag{3.9}$$

Thus, by (v) and (3.1), we deduce (3.5) from

$$s'(q) = \text{grad} (\varphi(qx(q)))(I + qA)^{-2}p > 0. \tag{3.10}$$

Using this theorem, we can easily claim that the function $F$ in (3.2) has the same characteristics as the function $f$ introduced in the previous section.

**Corollary 3.2.** Assume that (i)–(v) hold. Then, the function $F$ verifies all the hypotheses (a)–(c).

**Proof.** Assertions (a) and (b) directly follow from Theorem 3.1. Moreover, in view of (3.5), (c) is immediately obtained by $qF(q) = s(q)q^{1-\gamma}$, recalling that $\gamma < 1$ implies $q^{1-\gamma}$ strictly increasing.

Now, we are ready to state our main result about the convergence of the sequence $\{x_r\}$ defined as

$$x_{r+1} = (I + \varphi(x_r)A)^{-1}p, \quad r = 0, 1, \ldots, x_0 \geq 0. \tag{3.11}$$

**Theorem 3.3.** Assume that the hypotheses (i)–(v) hold. Then, the sequence $\{x_r\}$ converges to a solution $x^*$ of (1.2).

**Proof.** Let us start from an arbitrary $x_0 \geq 0$ and put

$$q_r = \varphi(x_r), \quad r = 0, 1, \ldots \tag{3.12}$$

From (3.11), we have that (3.12) can also be written as $q_{r+1} = F(q_r), r = 0, 1, \ldots$. Therefore, Corollary 3.2 assures that the sequence $\{q_r\}$ is convergent. Denoted by $q^*$ its limit, we set $x^* := (I + q^*A)^{-1}p$, and we obtain the statement by observing that

$$\lim_{r} x_{r+1} = \lim_{r} (I + \varphi(x_r)A)^{-1}p = \lim_{r} (I + q_r A)^{-1}p = (I + q^*A)^{-1}p = x^*. \tag{3.13}$$

**Remark 3.4.** Of course, in view of (3.7), any solution of (1.2) satisfies $\|x^*\|_{\infty} \leq \|p\|_{\infty}$, but nothing seems to imply its uniqueness. Anyway, we conjecture it, because in a large variety of
experimental tests (including also high-dimensional systems), we never found more than one solution. Of course, a sufficient condition for the uniqueness is \( \| \text{grad}(\varphi(x)) \| \| A \| \| p \|_\infty < 1 \), for any \( x \geq 0 \), which assures \( |F'(q)| < 1 \), for any \( q \geq 0 \), by virtue of (3.4) and (3.7). Nevertheless, as it can be observed in the next section, it seems that the solution is unique even if such a condition is largely not satisfied.

4. Numerical Experiments

In order to verify our theoretical results and to check the performances of the FP iteration scheme on problem (1.2), we carried out a large variety of numerical experiments. Here, the most significant are reported. From the previous section, it appears clear that the iterative scheme (3.11), which furnishes a vector at each iteration, requires the same computational effort as (3.12), which instead provides scalar values. Hence, for the sake of simplicity, all the numerical tests reported here, are referred to (3.12), that is, we test the iterative procedure:

\[
q_{r+1} = \varphi \left( (I + q_r A)^{-1} p \right) = F(q_r).
\] (4.1)

In particular, in all the section, we assume that the convergence is reached whenever \( |q_{r+1} - q_r| < \text{tol} \) and we put \( \text{tol} = 10^{-6} \).

Let us consider the following two problems:

\[
N = 3, \quad \varphi(x) = \| x \|_2, \quad A = \begin{pmatrix} 1 & -1 & 0 \\ -.5 & 1 & -.5 \\ -.5 & -1 & 2 \end{pmatrix}, \quad p = \begin{pmatrix} 5 \\ 1 \\ 10 \end{pmatrix}, \quad \gamma = 1; \quad (4.2)
\]

\[
N = 3, \quad \varphi(x) = \sqrt{x_1} \exp \left( \frac{x_1}{1 x_1 + 100 x_2 + x_3} \right), \quad x = (x_1, x_2, x_3)^T, \quad A = \begin{pmatrix} 10 & -10 & 0 \\ -10 & 15 & -.5 \\ 0 & -1 & 2 \end{pmatrix}, \quad p = \begin{pmatrix} .5 \\ 100 \\ 2 \end{pmatrix}, \quad \gamma = \frac{1}{2} \quad (4.3)
\]

which, as it can be easily checked, satisfy the hypotheses of Theorem 3.3. In Tables 1 and 2, we report the number of iterations we performed in order to compute the fixed point \( q^* \) of problems (4.2) and (4.3), respectively, for different choices of the starting value \( q_0 \). Moreover, we also report the number of iterations performed by the Matlab routine \texttt{fzero}. This is given by the sum of the number of iterations to find an interval containing the zero (first addendum in Table 1) and the number of zero-finding iterations (second addendum in Table 1).

From these two tests, we observe that, according to Theorem 3.3, the FP iteration method converges for any \( q_0 \) in both the cases. Of course, when we know the interval where a zero lies, \texttt{fzero} does converge and it requires much less iterations, but in many cases this information is unavailable and, see Table 2, \texttt{fzero} fails to provide a value. In fact, at the third line of Table 2, the symbol AB stays for “aborted” and means that the MATLAB routine does not succeed in finding an interval containing a sign change. The last line of Table 2...
Table 1

<table>
<thead>
<tr>
<th>FP iterations.</th>
<th>15</th>
<th>14</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_0 )</td>
<td>.2</td>
<td>.4</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>( \text{fzero} ) iterations.</td>
<td>21 + 4</td>
<td>19 + 4</td>
<td>16 + 4</td>
<td>12 + 6</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>FP iterations.</th>
<th>19</th>
<th>19</th>
<th>19</th>
<th>17</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_0 )</td>
<td>.05</td>
<td>.1</td>
<td>1</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>( \text{fzero} ) iterations.</td>
<td>AB</td>
<td>AB</td>
<td>AB</td>
<td>11</td>
<td>AB</td>
</tr>
<tr>
<td>FP + ( \text{fzero} ) iterations.</td>
<td>1 + 13</td>
<td>1 + 15</td>
<td>1 + 12</td>
<td>3 + 4</td>
<td>2 + 13</td>
</tr>
</tbody>
</table>

shows that the best results can be obtained by a combined solution strategy, which starts with a certain number of FP iterations (first addendum) and takes the final steps using \( \text{fzero} \) iterations (second addendum). At the fifth column, we take the right balance between FP and \( \text{fzero} \) iterations, while in the remaining columns, we displayed the minimum number of FP iterations which are necessary to get a value from \( \text{fzero} \). Thus, we can fruitfully exploit the global convergence of our FP iteration method in order to create more efficient combined strategy with \( \text{fzero} \). Of course, this could be done also by means of other well-known methods like Newton’s one.

In order to verify whether the convergence of the FP procedure could be ascribed to the Banach contraction theorem, we plotted the absolute value of the derivative of the function \( F \) in (4.1) (not reported) and we noted that the starting value in bold, that is, \( q_0 = .2, q_0 = .4 \) in Table 1 and \( q_0 = .05, q_0 = .1 \) in Table 2, belongs to intervals where \(|F'(q)| > 1\). Hence we can assert that, at least in these cases, the convergence of the FP method is not “helped” by the contraction principle.

As we already wrote in Remark 3.4, from these numerical experiments and many others, starting from very different values of \( q_0 \), the method seems to converge to the same fixed point. For example, in problems (4.2) and (4.3), we obtain \( q^* \approx 4.7954 \) and \( q^* \approx 2.6085 \), respectively, for a very large number of initial guesses (much larger than the ones reported in Tables 1 and 2). This numerical evidence let us conjecture that the assumptions in Theorem 3.3 are also sufficient to assure the uniqueness of the solution of problem (1.2).

In Figures 1 and 2, we plotted the first 100 elements of the error sequence \(|q_r - q^*|\), where \( q_r \) is computed by applying the FP iterations to the following two problems:

\[
N = 3, \quad \varphi(x) = x_1\sin^2\left(\frac{20x_1}{.1x_1 + x_2 + x_3}\right), \quad q^* = 1.1210, \\
A = \begin{pmatrix} 10 & -10 & 0 \\
-10 & 15 & -5 \\
0 & -1 & 2 \end{pmatrix}, \quad p = \begin{pmatrix} .5 \\
100 \\
2 \end{pmatrix}, \quad \gamma = 1; 
\]

(4.4)

\[
N = 1, \quad \varphi(x) = x^{3/2}, \quad q^* = 0.0898, \quad A = 100, \quad p = 2, \quad \gamma = \frac{3}{2}. 
\]

(4.5)

It can be easily seen that both the problems do not satisfy the hypotheses of Theorem 3.3. To be more precise, (4.4) does not fulfill (v), whereas (4.5) does not verify (iv), having \( \gamma > 1 \). The
figures clearly show that the method fails to converge, suggesting us that the assumptions of our main theorem are not too restrictive.

Moreover, we point out that, in our experience, some problems arising in the kinetic theory of dusty plasmas, are more cumbersome than (1.1), and therefore, the application of Newton or other available iterative procedures is not always simple, convenient, or possible. A comparison between FP and other iteration processes on problem of type (1.1) will be the subject of future investigations. Finally, in order to test the performance of our method for larger systems arising from the applications, we discretized a problem of type (1.1) with stepsizes $h$ on a sufficiently large interval $(0, T)$, with $T = Nh$. In this way, we get a nonlinear system of type (1.2), which has dimension $N$. By applying FP iteration method to such a system for different values of $N$, with $10^2 < N < 10^5$, we observe that the number of iterations does not vary with $N$ and so with $h$, but only with the choice of the starting point $q_0$. Moreover, we underline that in this case the linear systems arising from the FP iterations are tridiagonal and then very easy to be solved.
References


