DYNAMICAL SYSTEMS METHOD FOR SOLVING NONLINEAR EQUATIONS WITH MONOTONE OPERATORS

N. S. HOANG AND A. G. RAMM

Abstract. A version of the Dynamical Systems Method (DSM) for solving ill-posed nonlinear equations with monotone operators in a Hilbert space is studied in this paper. An a posteriori stopping rule, based on a discrepancy-type principle is proposed and justified mathematically. The results of two numerical experiments are presented. They show that the proposed version of DSM is numerically efficient. The numerical experiments consist of solving nonlinear integral equations.

1. Introduction

In this paper we study a Dynamical Systems Method (DSM) for solving the equation

\begin{equation}
F(u) = f,
\end{equation}

where \( F \) is a nonlinear twice Fréchet differentiable monotone operator in a real Hilbert space \( H \), and equation (1.1) is assumed solvable. Monotonicity means that

\begin{equation}
\langle F(u) - F(v), u - v \rangle \geq 0, \quad \forall u, v \in H.
\end{equation}

Here \( \langle \cdot, \cdot \rangle \) denotes the inner product in \( H \). It is known (see, e.g., [8]), that the set \( \mathcal{N} := \{ u : F(u) = f \} \) is closed and convex if \( F \) is monotone and continuous. A closed and convex set in a Hilbert space has a unique minimal-norm element. This element in \( \mathcal{N} \) we denote \( y, F(y) = f \). We assume that

\begin{equation}
\sup_{\|u-u_0\| \leq R} \|F^{(j)}(u)\| \leq M_j(u_0, R), \quad 0 \leq j \leq 2,
\end{equation}

where \( u_0 \in H \) is an element of \( H \), \( R > 0 \) is arbitrary, and \( f = F(y) \) is not known; but \( f_\delta \), the noisy data, are known and \( \|f_\delta - f\| \leq \delta \). If \( F(u) \) is not boundedly invertible, then solving for \( u \) given noisy data \( f_\delta \) is often (but not always) an ill-posed problem. When \( F \) is a linear bounded operator many methods for a stable solution of (1.1) were proposed (see [4], [5], [6], [7], [8] and the references therein). However, when \( F \) is nonlinear then the theory is less complete.

DSM for solving equation (1.1) was extensively studied in [8], [9], [10], [11], [12], [13], [14], [15]. In [8] the following version of the DSM for solving equation (1.1) was studied:

\begin{equation}
\dot{u}_\delta = -(F'(u_\delta) + a(t)I)^{-1}(F(u_\delta) + a(t)u_\delta - f_\delta), \quad u_\delta(0) = u_0.
\end{equation}

Received by the editor April 3, 2008, and, in revised form, January 17, 2009.

2000 Mathematics Subject Classification. Primary 65R30; Secondary 47J05, 47J06, 47J35.

Key words and phrases. Dynamical systems method (DSM), nonlinear operator equations, monotone operators, discrepancy principle.
Here $F$ is a monotone operator, and $a(t) > 0$ is a continuous function, defined for all $t \geq 0$, strictly monotonically decaying, $\lim_{t \to \infty} a(t) = 0$. These assumptions on $a(t)$ hold throughout the paper and are not repeated. Additional assumptions on $a(t)$ will appear later. Convergence of the above DSM was proved in [8] for any initial value $u_0$ with an a priori choice of stopping time $t_δ$, provided that $a(t)$ is suitably chosen.

The theory of monotone operators is presented in many books, e.g., in [1], [7], [16]. Most of the results of the theory of monotone operators, used in this paper, can be found in [8]. In [6] methods for solving nonlinear equations in a finite-dimensional space are discussed.

In this paper we propose and justify a stopping rule based on a discrepancy principle (DP) for the DSM (1.4). The main result of this paper is Theorem 3.1 in which a DP is formulated, the existence of the stopping time $t_δ$ is proved, and the convergence of the DSM with the proposed DP is justified under some natural assumptions apparently for the first time for a wide class of nonlinear equations with monotone operators.

These results are new from the theoretical point of view and very useful practically. The auxiliary results in our paper are also new and can be used in other problems of numerical analysis. These auxiliary results are formulated in Lemmas 2.2, 2.3, 2.7, 2.10, 2.11, and in the remarks. In particular, in Remark 3.3 we emphasize that the trajectory of the solution stays in a ball of a fixed radius $R$ for all $t \geq 0$.

In Section 4 the results of two numerical experiments are presented. In the second experiment we demonstrate numerically that our method for solving equation (1.1) can be used even for a wider class of equations than the basic Theorem 3.1 guarantees.

2. Auxiliary Results

Let us consider the following equation:

\[ F(V_{δ,a}) + aV_{δ,a} - f_δ = 0, \quad a > 0, \]

where $a = \text{const}$. It is known (see, e.g., [8]) that equation (2.1) with monotone continuous operator $F$ has a unique solution for any $f_δ \in H$.

Let us recall the following result from [8, p. 112]:

**Lemma 2.1.** Assume that equation (1.1) is solvable, $y$ is its minimal-norm solution, and assumptions (1.2) and (1.3) hold. Then

\[ \lim_{a \to 0} \| V_{0,a} - y \| = 0, \]

where $V_{0,a}$ solves (2.1) with $δ = 0$.

**Lemma 2.2.** If (1.2) holds and $F$ is continuous, then $\| V_{δ,a} \| = O(\frac{1}{a})$ as $a \to \infty$, and

\[ \lim_{a \to \infty} \| F(V_{δ,a}) - f_δ \| = \| F(0) - f_δ \|. \]

**Proof.** Rewrite (2.1) as

\[ F(V_{δ,a}) - F(0) + aV_{δ,a} + F(0) - f_δ = 0. \]
Multiply this equation by $V_{\delta,a}$, use inequality $\langle F(V_{\delta,a}) - F(0), V_{\delta,a} - 0 \rangle \geq 0$ from (1.2) and get 
$$a\|V_{\delta,a}\|^2 \leq \langle aV_{\delta,a} + F(V_{\delta,a}) - F(0), V_{\delta,a} \rangle = \langle f_\delta - F(0), V_{\delta,a} \rangle \leq \|f_\delta - F(0)\|\|V_{\delta,a}\|.$$ 
Therefore, $\|V_{\delta,a}\| = O(\frac{1}{\delta})$. This and the continuity of $F$ imply (2.2). □

Let $a = a(t)$, $0 < a(t) \searrow 0$, and assume $a \in C^1[0,\infty)$. Then the solution $V_\delta(t) := V_{\delta,a(t)}$ of (2.1) is a function of $t$. From the triangle inequality one gets
$$\|F(V_\delta(0)) - f_\delta\| \geq \|F(0) - f_\delta\| - \|F(V_\delta(0)) - F(0)\|.$$ 
From Lemma 2.2 it follows that for large $a(0)$ one has
$$\|F(V_\delta(0)) - F(0)\| \leq M_1\|V_\delta(0)\| = O\left(\frac{1}{a(0)}\right).$$ 
Therefore, if $\|F(0) - f_\delta\| > C\delta$, then $\|F(V_\delta(0)) - f_\delta\| \geq (C - \epsilon)\delta$, where $\epsilon > 0$ is sufficiently small, for sufficiently large $a(0) > 0$.

Below the words decreasing and increasing mean strictly decreasing and strictly increasing.

**Lemma 2.3.** Assume $\|F(0) - f_\delta\| > 0$. Let $0 < a(t) \searrow 0$, and let $F$ be monotone. Denote
$$\phi(t) := \|F(V_\delta(t)) - f_\delta\|, \quad \psi(t) := \|V_\delta(t)\|,$$ 
where $V_\delta(t)$ solves (2.1) with $a = a(t)$. Then $\phi(t)$ is decreasing, and $\psi(t)$ is increasing.

**Proof.** Since $\|F(0) - f_\delta\| > 0$, it follows that $\psi(t) \neq 0$, $\forall t \geq 0$. Note that $\phi(t) = a(t)\|V_\delta(t)\|$. One has
$$0 \leq \langle F(V_\delta(t_1)) - F(V_\delta(t_2)), V_\delta(t_1) - V_\delta(t_2) \rangle$$
$$= \langle -(a(t_1)V_\delta(t_1) + a(t_2)V_\delta(t_2)), V_\delta(t_1) - V_\delta(t_2) \rangle$$
$$= (a(t_1) + a(t_2))(V_\delta(t_1), V_\delta(t_2)) - a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2.$$ 
Thus,
$$0 \leq (a(t_1) + a(t_2))(V_\delta(t_1), V_\delta(t_2)) - a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2$$
$$\leq (a(t_1) + a(t_2))(\|V_\delta(t_1)\|^2 - a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2)$$
$$= (a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2)(\|V_\delta(t_2)\| - \|V_\delta(t_1)\|)$$
$$= (\phi(t_1) - \phi(t_2))(\psi(t_2) - \psi(t_1)).$$
If $\psi(t_2) > \psi(t_1)$, then (2.4) implies $\phi(t_1) \geq \phi(t_2)$, so
$$a(t_1)\psi(t_1) \geq a(t_2)\psi(t_2) > a(t_2)\psi(t_1).$$ 
Thus, if $\psi(t_2) > \psi(t_1)$, then $a(t_2) < a(t_1)$ and, therefore, $t_2 > t_1$, because $a(t)$ is decreasing.

Similarly, if $\psi(t_2) < \psi(t_1)$, then $\phi(t_1) < \phi(t_2)$. This implies $a(t_2) > a(t_1)$, so $t_2 < t_1$.

If $\psi(t_2) = \psi(t_1)$, then (2.3) implies
$$\|V_\delta(t_1)\|^2 \leq \langle V_\delta(t_1), V_\delta(t_2) \rangle \leq \|V_\delta(t_1)\|\|V_\delta(t_2)\| = \|V_\delta(t_1)\|^2.$$ 
This implies $V_\delta(t_1) = V_\delta(t_2)$, and then $a(t_1) = a(t_2)$. Hence, $t_1 = t_2$, because $a(t)$ is decreasing.

Therefore, $\phi(t)$ is decreasing and $\psi(t)$ is increasing. □
Lemma 2.4. Suppose that \( \|F(0) - f_\delta\| > C\delta \), \( C > 1 \), and \( a(0) \) is sufficiently large. Then, there exists a unique \( t_1 > 0 \) such that \( \|F(V_\delta(t_1)) - f_\delta\| = C\delta \).

Proof. The uniqueness of \( t_1 \) follows from Lemma 2.3. We have \( F(y) = f \), and

\[
0 = \langle F(V_\delta) + aV_\delta - f_\delta, F(V_\delta) - f_\delta \rangle = \|F(V_\delta) - f_\delta\|^2 + a\langle V_\delta - y, F(V_\delta) - f_\delta \rangle + a\langle y, F(V_\delta) - f_\delta \rangle + a\langle y, y, F(V_\delta) - f_\delta \rangle \\
\geq \|F(V_\delta) - f_\delta\|^2 + a\langle V_\delta - y, f - f_\delta \rangle + a\langle y, F(V_\delta) - f_\delta \rangle.
\]

(2.5)

Here the inequality \( \langle V_\delta - y, F(V_\delta) - F(y) \rangle \geq 0 \) was used. Therefore,

\[
\|F(V_\delta) - f_\delta\|^2 \leq -a\langle V_\delta - y, f - f_\delta \rangle - a\langle y, F(V_\delta) - f_\delta \rangle \\
\leq a\|V_\delta - y\|\|f - f_\delta\| + a\|y\|\|F(V_\delta) - f_\delta\| \\
\leq a\delta\|V_\delta - y\| + a\|y\|\|F(V_\delta) - f_\delta\|.
\]

On the other hand, we have

\[
0 = \langle F(V_\delta) - F(y) + aV_\delta + f - f_\delta, V_\delta - y \rangle \\
= \langle F(V_\delta) - F(y), V_\delta - y \rangle + a\|V_\delta - y\|^2 + a\langle y, V_\delta - y \rangle + (f - f_\delta, V_\delta - y),
\]

where the inequality \( \langle V_\delta - y, F(V_\delta) - F(y) \rangle \geq 0 \) was used. Therefore,

\[
a\|V_\delta - y\|^2 \leq a\|y\|\|V_\delta - y\| + \delta\|V_\delta - y\|.
\]

This implies

\[
a\|V_\delta - y\| \leq a\|y\| + \delta.
\]

(2.6)

From (2.5) and (2.6), and an elementary inequality \( ab \leq \epsilon a^2 + \frac{b^2}{4\epsilon}, \forall \epsilon > 0 \), one gets

\[
\|F(V_\delta) - f_\delta\|^2 \leq \delta^2 + a\|y\|\delta + a\|y\|\|F(V_\delta) - f_\delta\| \\
\leq \delta^2 + a\|y\|\delta + \epsilon\|F(V_\delta) - f_\delta\|^2 + \frac{1}{4\epsilon}\|y\|^2,
\]

(2.7)

where \( \epsilon > 0 \) is fixed, independent of \( t \), and can be chosen arbitrarily small. Let \( t \to \infty \) and \( a = a(t) \uparrow 0 \). Then (2.7) implies \( \limsup_{t \to \infty} (1 - \epsilon)\|F(V_\delta) - f_\delta\|^2 \leq \delta^2 \). This, the continuity of \( F \), the continuity of \( V_\delta(t) \) on \([0, \infty)\), and the assumption \( \|F(0) - f_\delta\| > C\delta \), where \( C > 1 \), imply that equation \( \|F(V_\delta(t)) - f_\delta\| = C\delta \) must have a solution \( t_1 > 0 \).

\[ \square \]

Remark 2.5. Let \( V := V_\delta(t) \big|_{\delta=0} \), so \( F(V) + a(t)V - f = 0 \). Let \( y \) be the minimal-norm solution to \( F(u) = f \). We claim that

\[
\|V_\delta - V\| \leq \frac{\delta}{a}.
\]

(2.8)

Indeed, from (2.1) one gets

\[
F(V_\delta) - F(V) + a(V_\delta - V) = f - f_\delta.
\]
Multiply this equality by \((V_\delta - V)\) and use (1.2) to obtain
\[
\delta\|V_\delta - V\| \geq \langle f - f_\delta, V_\delta - V \rangle \\
= \langle F(V_\delta) - F(V) + a(V_\delta - V), V_\delta - V \rangle \\
\geq a\|V_\delta - V\|^2.
\]
This implies (2.8).
Similarly, from the equation
\[
F(V) + aV - F(y) = 0,
\]
one can derive that
\[
\|V\| \leq \|y\|.
\]
From (2.8) and (2.9), one gets the following estimate:
\[
\|V_\delta\| \leq \|V\| + \frac{\delta}{\alpha} \leq \|y\| + \frac{\delta}{\alpha}.
\]

Let us recall the following lemma, which is basic in our proofs.

**Lemma 2.6** (8, p. 97). Let \(\alpha(t), \beta(t), \gamma(t)\) be continuous nonnegative functions on \([\tau_0, \infty)\), \(\tau_0 \geq 0\) is a fixed number. If there exists a function \(\mu := \mu(t)\),
\[
\mu \in C^1[\tau_0, \infty), \quad \mu(t) > 0, \quad \lim_{t \to \infty} \mu(t) = \infty,
\]
such that
\[
0 \leq \alpha(t) \leq \frac{\mu(t)}{2} \left[ \gamma - \frac{\dot{\mu}(t)}{\mu(t)} \right], \quad \dot{\mu} := \frac{d\mu}{dt},
\]
\[
\beta(t) \leq \frac{1}{2\mu(t)} \left[ \gamma - \frac{\dot{\mu}(t)}{\mu(t)} \right],
\]
\[
\mu(\tau_0)g(\tau_0) < 1,
\]
and \(g(t) \geq 0\) satisfies the inequality
\[
g(t) \leq -\gamma(t)g(t) + \alpha(t)g^2(t) + \beta(t), \quad t \geq \tau_0,
\]
then
\[
0 \leq g(t) < \frac{1}{\mu(t)} \to 0, \quad \text{as} \quad t \to \infty.
\]
If inequalities (2.11) - (2.13) hold on an interval \([\tau_0, T]\), then \(g(t)\), the solution to inequality (2.14), exists on this interval and inequality (2.15) holds on \([\tau_0, T]\).

**Lemma 2.7.** Suppose \(M_1, c_0, \text{ and } c_1\) are positive constants and \(0 \neq y \in H\). Then there exist \(\lambda > 0\) and a function \(a(t) \in C^1[0, \infty), \ 0 < a(t) \searrow 0\), such that the following conditions hold:
\[
\frac{M_1}{\|y\|} \leq \lambda,
\]
\[
\frac{c_0}{a(t)} \leq \frac{\lambda}{2\alpha(t)} \left[ 1 - \frac{|\dot{a}(t)|}{a(t)} \right],
\]
\[
\frac{c_1|\dot{a}(t)|}{a(t)} \leq \frac{a(t)}{2\lambda} \left[ 1 - \frac{|\dot{a}(t)|}{a(t)} \right],
\]
\[
\|F(0) - f_\delta\| \leq \frac{a^2(0)}{\lambda}.
\]
Proof. Take
\begin{equation}
(2.20) \quad a(t) = \frac{d}{(c + t)^b}, \quad 0 < b \leq 1, \quad c \geq \max(2b, 1).
\end{equation}

Note that \(|\dot{a}| = -\ddot{a}|. We have
\begin{equation}
(2.21) \quad \frac{\dot{a}(t)}{a(t)} = \frac{b}{c + t} \leq \frac{b}{c} \leq \frac{1}{2}, \quad \forall t \geq 0.
\end{equation}

Hence,
\begin{equation}
(2.22) \quad \frac{1}{2} \leq 1 - \frac{\dot{a}(t)}{a(t)}, \quad \forall t \geq 0.
\end{equation}

Take
\begin{equation}
(2.23) \quad \lambda \geq \frac{M_1}{\|y\|}.
\end{equation}

Then \((2.16)\) is satisfied.

Choose \(d\) such that
\begin{equation}
(2.24) \quad d \geq \max \left( \sqrt{\frac{c^2 \lambda}{2} \|F(0) - f_\delta\|}, 4b c_1 \right).
\end{equation}

From equality \((2.20)\) and inequality \((2.24)\) one gets
\begin{equation}
(2.25) \quad \frac{\dot{a}(t)}{a^2(t)} = \frac{b}{d(c + t)^{1-b}} \leq \frac{b}{d} \leq \frac{1}{4 \lambda c_1}, \quad \forall t \geq 0.
\end{equation}

This and inequality \((2.21)\) imply inequality \((2.18)\). It follows from inequality \((2.24)\) that
\begin{equation}
(2.26) \quad \|F(0) - f_\delta\| \leq \frac{d^2}{c^2 \lambda} = \frac{a^2(0)}{\lambda}.
\end{equation}

Thus, inequality \((2.19)\) is satisfied.

Choose \(\kappa \geq 1\) such that
\begin{equation}
(2.27) \quad \kappa > \max \left( \frac{4c_0}{\lambda}, 1 \right).
\end{equation}

Define
\begin{equation}
(2.28) \quad \nu(t) := \kappa a(t), \quad \lambda_\kappa := \kappa \lambda.
\end{equation}

Note that inequalities \((2.10), (2.15), (2.19)\) and \((2.21)\) still hold for \(a(t) = \nu(t)\) and \(\lambda = \lambda_\kappa\).

Using the inequalities \((2.21)\) and \(c \geq 1\) and the definition \((2.28)\), one obtains
\begin{equation}
(2.29) \quad \frac{c_0}{\nu(t)} \leq \frac{\lambda_\kappa}{4 \nu(t)} \leq \frac{\lambda_\kappa}{2 \nu(t)} \left[ 1 - \frac{\dot{\nu}}{\nu} \right].
\end{equation}

Thus, one can replace the function \(a(t)\) by \(\nu(t) = \kappa a(t)\) and \(\lambda\) by \(\lambda = \lambda_\kappa\) to satisfy inequalities \((2.16) - (2.19)\). \(\square\)

Remark 2.8. In the proof of Lemma 2.7 \(a(0)\) and \(\lambda\) can be chosen so that \(\frac{a(0)}{\lambda}\) is uniformly bounded as \(\delta \to 0\) regardless of the rate of growth of the constant \(M_1 = M_1(R)\) from formula \((1.3)\) when \(R \to \infty\), i.e., regardless of the strength of the nonlinearity \(F(u)\).
Indeed, to satisfy (2.23) one can choose \( \lambda = \frac{M}{\|s\|} \). To satisfy (2.24) one can choose

\[
d = \max \left( \sqrt{c^2} \lambda \| f_\delta - F(0) \|, 4b \lambda c_1 \right) \leq \max \left( \sqrt{c^2} \lambda (\| f - F(0) \| + 1), 4b \lambda c_1 \right),
\]

where we have assumed, without loss of generality, that \( 0 < \delta < 1 \). With this choice of \( d \) and \( \lambda \), the ratio \( \frac{a(0)}{\lambda} \) is bounded uniformly with respect to \( \delta \in (0, 1) \) and does not depend on \( R \).

Indeed, with the above choice one has \( \frac{a(0)}{\lambda} = \frac{d}{\sqrt{c^2}} \leq \tilde{c}(1 + \sqrt{\lambda^{-1}}) \leq \tilde{c} \), where \( \tilde{c} > 0 \) is a constant independent of \( \delta \), and one can assume that \( \lambda \geq 1 \) without loss of generality.

This remark is used in Remark 3.3, where we prove that the trajectory of \( u_\delta(t) \), defined by (2.11), stays in a ball \( B(u_0, R) \) for all \( 0 \leq t \leq t_\delta \), where the number \( t_\delta \) is defined by formula (3.1) (see below), and \( R > 0 \) is sufficiently large. An upper bound on \( R \) is given in Remark 3.3.

**Remark 2.9.** It is easy to choose \( u_0 \in H \) such that

\[
g_0 := \| u_0 - V_\delta(0) \| \leq \frac{\| F(0) - f_\delta \|}{a(0)}.
\]

Indeed, if, for example, \( u_0 = 0 \), then by Lemmas 2.22 and 2.23 one gets

\[
g_0 = \| V_\delta(0) \| = \frac{a(0)\| V_\delta(0) \|}{a(0)} \leq \frac{\| F(0) - f_\delta \|}{a(0)}.
\]

If (2.19) and (2.30) hold, then \( g_0 \leq \frac{a(0)}{\lambda} \). Inequality (2.30) also holds if \( \| u_0 - V_\delta(0) \| \) is sufficiently small.

**Lemma 2.10.** Let \( p, b \) and \( c \) be positive constants. Then

\[
(p - \frac{b}{c}) \int_0^t \frac{e^{ps}}{(s + c)^b} ds < \frac{e^{pt}}{(c + t)^b}, \quad \forall c, b > 0, \quad t > 0.
\]

**Proof.** One has

\[
\frac{d}{dt} \left( \frac{e^{pt}}{(c + t)^b} \right) = \frac{pe^{pt}}{(c + t)^{b+1}} - \frac{be^{pt}}{(c + t)^b} \geq \left( p - \frac{b}{c} \right) \frac{e^{pt}}{(c + t)^b}, \quad t \geq 0.
\]

Therefore,

\[
\left( p - \frac{b}{c} \right) \int_0^t \frac{e^{ps}}{(s + c)^b} ds \leq \int_0^t \frac{d}{ds} \left( \frac{e^{ps}}{(s + c)^b} \right) ds \leq \frac{e^{pt}}{(c + t)^b} - \frac{1}{c^b} \leq \frac{e^{pt}}{(c + t)^b}.
\]

Lemma 2.10 is proved.

**Lemma 2.11.** Let \( a(t) = \frac{d}{(c + t)^b} \) where \( d, c, b > 0, \ c \geq 6b \). One has

\[
e^{-\frac{t}{2}} \int_0^t e^{\frac{s}{2}} |\dot{u}(s)||V_\delta(s)||ds \leq \frac{1}{2} a(t)||V_\delta(t)||, \quad t \geq 0.
\]
Proof. Let $p = \frac{1}{2}$ in Lemma 2.10 Then

$$(2.33) \quad \left(1 - \frac{b}{c}\right)\int_0^t e^{\frac{s}{(s+c)^b}} ds < \frac{c}{(c+t)^b}, \quad \forall c, b \geq 0.$$  

Since $c \geq 6b$ or $\frac{2b}{c} \leq \frac{1}{2}$, one has

$$\frac{1}{2} - \frac{b}{c} \geq \frac{2b}{c + s}, \quad s \geq 0.$$  

This implies

$$(2.34) \quad a(s)\left(\frac{1}{2} - \frac{b}{c}\right) = \frac{d}{(c+s)^b}\left(\frac{1}{2} - \frac{b}{c}\right) \geq \frac{2db}{(c+s)^{b+1}} = 2|\dot{a}(s)|, \quad s \geq 0.$$  

Multiplying (2.34) by $e^{\frac{s}{(s+c)^b}}$, integrating from 0 to $t$, using inequality (2.33) and the fact that $\|V_\delta(s)\|$ is nondecreasing, one gets

$$e^{\frac{s}{(s+c)^b}}\|V_\delta(t)\| > \int_0^t e^{\frac{s}{(s+c)^b}}\|V_\delta(t)\|a(s)\left(\frac{1}{2} - \frac{b}{c}\right) ds \geq 2\int_0^t e^{\frac{s}{(s+c)^b}}\|\dot{a}(s)\|V_\delta(s) ds, \quad t \geq 0.$$  

This implies inequality (2.32). Lemma 2.11 is proved.  

3. MAIN RESULT

Denote

$$A := F'(u_\delta(t)), \quad A_\delta := A + a I,$$

where $I$ is the identity operator, and $u_\delta(t)$ solves the following Cauchy problem:

$$(3.1) \quad \dot{u}_\delta = -A^{-1}\partial_t[F(u_\delta) + a(t)u_\delta - f_\delta], \quad u_\delta(0) = u_0.$$  

We assume below that $\|F(u_0) - f_\delta\| > C_1\delta^\zeta$, where $C_1 > 1$ and $\zeta \in (0, 1]$ are some constants. We also assume, without loss of generality, that $\delta \in (0, 1)$.

Assume that equation $F(u) = f$ has a solution, possibly nonunique, and $y$ is the minimal norm solution to this equation. Let $f$ be unknown, but $f_\delta$ be given, $\|f_\delta - f\| \leq \delta$.

**Theorem 3.1.** Assume $a(t) = \frac{d}{(c+t)^b}$, where $b \in (0, 1], c, d > 0$ are constants, $c > 6b$, and $d$ is sufficiently large so that inequalities (2.17)–(2.19) hold. Assume that $F : H \to H$ is a monotone operator, twice Fréchet differentiable, $\sup_{u \in B(u_0, R)} \|F^{(j)}(u)\| \leq M_j(u_0, R), 0 \leq j \leq 2, B(u_0, R) := \{u : \|u - u_0\| \leq R\}$, $u_0$ is an element of $H$, satisfying inequality (2.30) and

$$(3.2) \quad \|F(u_0) + a(0)u_0 - f_\delta\| \leq \frac{1}{4}a(0)\|V_\delta(0)\|,$$

where $V_\delta(t) := V_{\delta, a(t)}$ solves (2.1) with $a = a(t)$. Then the solution $u_\delta(t)$ to problem (3.1) exists on an interval $[0, T_\delta]$, $\lim_{t \to 0} T_\delta = \infty$, and there exists a unique $t_\delta \in (0, T_\delta)$ such that $\lim_{t \to 0} t_\delta = \infty$ and

$$(3.3) \quad \|F(u_\delta(t)) - f_\delta\| = C_1\delta^\zeta, \quad \|F(u_\delta(t)) - f_\delta\| > C_1\delta^\zeta, \quad \forall t \in [0, t_\delta],$$

where $C_1 > 1$ and $0 < \zeta \leq 1$. If $\zeta \in (0, 1)$ and $t_\delta$ satisfies (3.3), then

$$(3.4) \quad \lim_{\delta \to 0} \|u_\delta(t_\delta) - y\| = 0.$$
Remark 3.2. One can choose $u_0$ satisfying inequalities (2.30) and (3.2) (see also (3.34) below). Indeed, if $u_0$ is a sufficiently close approximation to $V_0(0)$, the solution to equation (2.1), then inequalities (2.30) and (3.2) are satisfied. Note that inequality (3.2) is a sufficient condition for (3.35) to hold. In our proof inequality (3.35) is used at $t = t_\delta$. The stopping time $t_\delta$ is often sufficiently large for the quantity $e^{-\frac{t_\delta}{h_0}}h_0$ to be small. In this case inequality (3.35) with $t = t_\delta$ is satisfied for a wide range of $u_0$. For example, in our numerical experiment in Section 4 the method converged rapidly when $u_0 = 0$.

Condition $c > 6b$ is used in the proof of Lemma 2.11.

Proof of Theorem 3.1. Denote

$$C := \frac{C_1 + 1}{2}.$$  \hspace{1cm} (3.5)

Let

$$w := u_\delta - V_\delta, \quad g(t) := \|w\|.$$  \hspace{1cm} (3.6)

One has

$$\dot{w} = -\dot{V}_\delta - A^{-1}_{a(t)}[F(u_\delta) - F(V_\delta) + a(t)w].$$  \hspace{1cm} (3.7)

We use Taylor’s formula and get

$$F(u_\delta) - F(V_\delta) + aw = A_w + K, \quad \|K\| \leq \frac{M_2}{2}\|w\|^2,$$  \hspace{1cm} (3.8)

where $K := F(u_\delta) - F(V_\delta) - A_w$, and $M_2$ is the constant from the estimate (1.3). Multiplying (3.6) by $w$ and using (3.7) one gets

$$\dot{g} \leq -g^2 + \frac{M_2}{2}a^{-1}_{a(t)}\|g\|^2 + \|\dot{V}_\delta\|g.$$  \hspace{1cm} (3.9)

Let $t_0$ be such that

$$\frac{\delta}{a(t_0)} = \frac{1}{C - 1}\|y\|, \quad C > 1.$$  \hspace{1cm} (3.10)

This $t_0$ exists and is unique since $a(t) > 0$ monotonically decays to 0 as $t \to \infty$.

Since $a(t) > 0$ monotonically decays, one has

$$\frac{\delta}{a(t)} \leq \frac{1}{C - 1}\|y\|, \quad 0 \leq t \leq t_0.$$  \hspace{1cm} (3.11)

By Lemma 2.3 there exists $t_1$ such that

$$\|F(V_\delta(t_1)) - f_\delta\| = C\delta, \quad F(V_\delta(t_1)) + a(t_1)V_\delta(t_1) - f_\delta = 0.$$  \hspace{1cm} (3.12)

We claim that $t_1 \in [0, t_0]$.

Indeed, from (2.4) and (2.10) one gets

$$C\delta = a(t_1)\|V_\delta(t_1)\| \leq a(t_1)\left(\|y\| + \frac{\delta}{a(t_1)}\right) = a(t_1)\|y\| + \delta, \quad C > 1,$$

so

$$\delta \leq \frac{a(t_1)\|y\|}{C - 1}.$$  \hspace{1cm} (3.13)

Thus,

$$\frac{\delta}{a(t_1)} \leq \frac{\|y\|}{C - 1} = \frac{\delta}{a(t_0)}.$$  \hspace{1cm} (3.14)

Since $a(t) \not\leq 0$, one has $t_1 \leq t_0$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Differentiating both sides of (2.12) with respect to \( t \), one obtains
\[
A_{a(t)} V_\delta = -\dot{a} V_\delta.
\]
This implies
\[
\|\dot{V}_\delta\| \leq |\dot{a}| \|A_{a(t)}^{-1} V_\delta\| \leq \frac{|\dot{a}|}{a} \|V_\delta\| \leq \frac{|\dot{a}|}{a} \left(\|y\| + \frac{\delta}{a}\right)
\]
(3.12)
\[
\leq \frac{|\dot{a}|}{a} \|y\| \left(1 + \frac{1}{C-1}\right), \quad \forall t \leq t_0.
\]
Since \( g \geq 0 \), inequalities (3.8) and (3.12) imply
\[
\dot{g} \leq -g(t) + \frac{c_0}{a(t)} g^2 + \frac{|\dot{a}|}{a(t)} c_1, \quad c_0 = \frac{M_2}{2}, \quad c_1 = \|y\| \left(1 + \frac{1}{C-1}\right).
\]
(3.13)
Here we have used the estimate
\[
\|A_{a(t)}^{-1}\| \leq \frac{1}{a}
\]
and the relations
\[
A_a := \tilde{F}'(u) + aI, \quad \tilde{F}'(u) := A \geq 0.
\]
Inequality (3.13) is of the type (2.14) with
\[
\gamma(t) = 1, \quad \alpha(t) = \frac{c_0}{a(t)}, \quad \beta(t) = c_1 \frac{|\dot{a}|}{a(t)}.
\]
Let us check assumptions (2.11)–(2.13). Take
\[
\mu(t) = \frac{\lambda}{a(t)},
\]
where \( \lambda = const > 0 \) and satisfies conditions (2.11)–(2.13) in Lemma 2.7. Since \( u_0 \) satisfies inequality (2.30), one gets \( g(0) \leq \frac{a(0)}{\lambda} \), by Remark 2.9. This, inequalities (2.11)–(2.13), and Lemma 2.6 yield
\[
g(t) < \frac{a(t)}{\lambda}, \quad \forall t \leq t_0, \quad g(t) := \|u_\delta(t) - V_\delta(t)\|.
\]
(3.14)
Therefore,
\[
\|F(u_\delta(t)) - f_\delta\| \leq \|F(u_\delta(t)) - F(V_\delta(t))\| + \|F(V_\delta(t)) - f_\delta\|
\]
(3.15)
\[
\leq M_1 g(t) + \|F(V_\delta(t)) - f_\delta\|
\]
\[
\leq \frac{M_1 a(t)}{\lambda} + \|F(V_\delta(t)) - f_\delta\|, \quad \forall t \leq t_0.
\]
It is proved in Section 2 Lemma 2.6 that \( \|F(V_\delta(t)) - f_\delta\| \) is decreasing. Since \( t_1 \leq t_0 \), one gets
\[
\|F(V_\delta(t_0)) - f_\delta\| \leq \|F(V_\delta(t_1)) - f_\delta\| = C_\delta.
\]
(3.16)
This, inequality (3.15), the inequality \( \frac{M_1}{\lambda} \leq \|y\| \) (see (2.23)), the relation (3.9), and the definition \( C_1 = 2C - 1 \) (see (3.5)), imply
\[
\|F(u_\delta(t_0)) - f_\delta\| \leq \frac{M_1 a(t_0)}{\lambda} + C_\delta 
\]
(3.17)
\[
\leq \frac{M_1 \delta (C - 1)}{\lambda \|y\|} + C_\delta \leq (2C - 1) \delta = C_1 \delta.
Thus, if
\[ \| F(u_\delta(0)) - f_\delta \| > C_1 \delta^\gamma, \quad 0 < \gamma \leq 1, \]
then, by the continuity of the function \( t \to \| F(u_\delta(t)) - f_\delta \| \) on \( [0, \infty) \), there exists \( t_\delta \in (0, t_0) \) such that
\[ (3.18) \quad \| F(u_\delta(t_\delta)) - f_\delta \| = C_1 \delta^\gamma \]
for any given \( \gamma \in (0, 1] \), and any fixed \( C_1 > 1 \).

Let us prove (3.18).

From (3.15) with \( t = t_\delta \), and from (2.10), one gets
\[ C_1 \delta^\gamma \leq M_1 \frac{a(t_\delta)}{\lambda} + a(t_\delta)\| V_\delta(t_\delta) \| \]
\[ \leq M_1 \frac{a(t_\delta)}{\lambda} + \| y \| a(t_\delta) + \delta. \]

Thus, for sufficiently small \( \delta \), one gets
\[ \tilde{C} \delta^\gamma \leq a(t_\delta) \left( \frac{M_1}{\lambda} + \| y \| \right), \quad \tilde{C} > 0, \]
where \( \tilde{C} < C_1 \) is a constant. Therefore,
\[ (3.19) \quad \lim_{\delta \to 0} \frac{\delta}{a(t_\delta)} \leq \lim_{\delta \to 0} \frac{\delta^{1-\gamma}}{\tilde{C}} \left( \frac{M_1}{\lambda} + \| y \| \right) = 0, \quad 0 < \zeta < 1. \]

We claim that
\[ (3.20) \quad \lim_{\delta \to 0} t_\delta = \infty. \]

Let us prove (3.20). Using (5.1), one obtains
\[ \frac{d}{dt} (F(u_\delta) + au_\delta - f_\delta) = A_\delta \dot{u}_\delta + \dot{a} u_\delta = - (F(u_\delta) + au_\delta - f_\delta) + \dot{a} u_\delta. \]
This and (2.1) imply
\[ (3.21) \quad \frac{d}{dt} [F(u_\delta) - F(V_\delta) + a(u_\delta - V_\delta)] = - [F(u_\delta) - F(V_\delta) + a(u_\delta - V_\delta)] + \dot{a} u_\delta. \]

Denote \( v := v(t) := F(u_\delta(t)) - F(V_\delta(t)) + a(t)(u_\delta(t) - V_\delta(t)), \quad h := h(t) := \| v \|. \)
Multiplying (3.21) by \( v \), one obtains
\[ \dot{h} h = - h^2 + \langle v, \dot{a}(u_\delta - V_\delta) \rangle + \dot{a} \langle v, V_\delta \rangle \]
\[ \leq - h^2 + h |\dot{a}| \| u_\delta - V_\delta \| + |\dot{a}| \| V_\delta \|, \quad h \geq 0. \]

Thus,
\[ (3.23) \quad \dot{h} \leq - h + |\dot{a}| \| u_\delta - V_\delta \| + |\dot{a}| \| V_\delta \|. \]

Since \( \langle F(u_\delta) - F(V_\delta), u_\delta - V_\delta \rangle \geq 0 \), one obtains from the two equations,
\[ \langle v, u_\delta - V_\delta \rangle = \langle F(u_\delta) - F(V_\delta) + a(t)(u_\delta - V_\delta), u_\delta - V_\delta \rangle \]
and
\[ \langle v, F(u_\delta) - F(V_\delta) \rangle = \| F(u_\delta) - F(V_\delta) \|^2 + a(t) \langle u_\delta - V_\delta, F(u_\delta) - F(V_\delta) \rangle, \]
the following two inequalities:
\[ (3.24) \quad a \| u_\delta - V_\delta \|^2 \leq \langle v, u_\delta - V_\delta \rangle \leq \langle u_\delta - V_\delta, h \rangle. \]
and
\[ (3.25) \quad \| F(u_\delta) - F(V_\delta) \|^2 \leq \langle v, F(u_\delta) - F(V_\delta) \rangle \leq h \| F(u_\delta) - F(V_\delta) \|. \]

Inequalities (3.24) and (3.25) imply
\[ (3.26) \quad a \| u_\delta - V_\delta \| \leq h, \quad \| F(u_\delta) - F(V_\delta) \| \leq h. \]

Inequalities (3.23) and (3.26) imply
\[ (3.27) \quad \dot{h} \leq -h\left(1 - \frac{\| \dot{a} \|}{a}\right) + |\dot{a}||V_\delta|. \]

Since \(1 - \frac{\| \dot{a} \|}{a} \geq \frac{1}{2}\) because \(c \geq 2b\), inequality (3.27) holds if
\[ (3.28) \quad \dot{h} \leq -\frac{1}{2}h + |\dot{a}||V_\delta|. \]

Inequality (3.28) implies
\[ (3.29) \quad h(t) \leq h(0)e^{-\frac{t}{2}} + e^{-\frac{t}{2}}\int_0^t e^\frac{s}{2} |\dot{a}||V_\delta| ds. \]

From (3.29) and (3.26), one gets
\[ (3.30) \quad \| F(u_\delta(t)) - F(V_\delta(t)) \| \leq h(0)e^{-\frac{t}{2}} + e^{-\frac{t}{2}}\int_0^t e^\frac{s}{2} |\dot{a}||V_\delta| ds. \]

Therefore,
\[ (3.31) \quad \| F(u_\delta(t)) - f_\delta \| \geq \| F(V_\delta(t)) - f_\delta \| - \| F(V_\delta(t)) - F(u_\delta(t)) \| \geq a(t)\|V_\delta(t)\| - h(0)e^{-\frac{t}{2}} - e^{-\frac{t}{2}}\int_0^t e^\frac{s}{2} |\dot{a}||V_\delta| ds. \]

From the results in Section 2 (see Lemma 2.11), it follows that there exists an \(a(t)\) such that
\[ (3.32) \quad \frac{1}{2}a(t)\|V_\delta(t)\| \geq e^{-\frac{t}{2}}\int_0^t e^\frac{s}{2} |\dot{a}||V_\delta(s)| ds. \]

For example, one can choose
\[ (3.33) \quad a(t) = \frac{d}{(e + t)^b}, \quad 6b < c, \]

where \(d, c, b > 0\). Moreover, one can always choose \(u_0\) such that
\[ (3.34) \quad h(0) = \| F(u_0) + a(0)u_0 - f_\delta \| \leq \frac{1}{4}a(0)\|V_\delta(0)\|, \]

because the equation \( F(u_0) + a(0)u_0 - f_\delta = 0 \) is solvable. If (3.34) holds, then
\[ \dot{h}(0)e^{-\frac{t}{2}} \leq \frac{1}{4}a(0)\|V_\delta(0)\|e^{-\frac{t}{2}}, \quad t \geq 0. \]

If \(2b < c\), then (3.33) implies
\[ e^{-\frac{t}{2}}a(0) \leq a(t). \]

Therefore,
\[ (3.35) \quad e^{-\frac{t}{2}}h(0) \leq \frac{1}{4}a(t)\|V_\delta(0)\| \leq \frac{1}{4}a(t)\|V_\delta(t)\|, \quad t \geq 0, \]
Moreover, from (3.36) and Lemma 2.1, one obtains (3.4). Theorem 3.1 is proved. □

Remark 3.3. The trajectory $u_\delta(t)$ remains in the ball $B(u_0, R) := \{ u : \| u - u_0 \| < R \}$ for all $t \leq t_\delta$, where $R$ does not depend on $\delta$ as $\delta \to 0$. Indeed, estimates (3.34), (2.10) and (3.10) imply

$$
\| u_\delta(t) - y \| \leq \| u_\delta(t) - V_\delta(t) \| + \| V(t) - V_\delta(t) \| + \| V(t) - y \|
$$

(3.37)

$$
\leq \frac{a(t_\delta)}{\lambda} + \frac{\delta}{a(t_\delta)} + \| V(t) - y \|.
$$

Note that $V(t_\delta) = V_0, a(t_\delta)$ (see equation (2.1)). From (3.14), (3.20), inequality (3.36) and Lemma 2.3 in Section 2. From (3.18) and (3.31)–(3.35), one gets

$$
\lim_{\delta \to 0} a(t_\delta)\| V_\delta(t_\delta) \| \leq \lim_{\delta \to 0} 4C_1\delta = 0.
$$

Since $\| V_\delta(t) \|$ increases (see Lemma 2.3), the above formula implies $\lim_{\delta \to 0} a(t_\delta) = 0$. Since $0 < a(t) \leq 0$, it follows that $\lim_{\delta \to 0} t_\delta = \infty$, i.e., (3.20) holds.

It is now easy to finish the proof of the Theorem 3.1.

From the triangle inequality and inequalities (3.14) and (2.8) one obtains

$$
\| u_\delta(t_\delta) - y \| \leq \| u_\delta(t_\delta) - V_\delta(t_\delta) \| + \| V(t_\delta) - V_\delta(t_\delta) \| + \| V(t_\delta) - y \|
$$

(3.36)

$$
\leq \frac{a(t_\delta)}{\lambda} + \frac{\delta}{a(t_\delta)} + \| V(t_\delta) - y \|.
$$

Here we have used the fact that $a(t) \geq 0$ (see the proof of Theorem 3.1). Since one can choose $a(t)$ and $\lambda$ so that $\frac{a(t)}{\lambda}$ is uniformly bounded as $\delta \to 0$ and regardless of the growth of $M_1$ (see Remark 2.8) one concludes that $R$ can be chosen independent of $\delta$ and $M_1$.

4. NUMERICAL EXPERIMENTS

4.1. An experiment with an operator defined on $H = L^2[0, 1]$. Let us do a numerical experiment solving nonlinear equation (1.1) with

$$
F(u) := B(u) + (\arctan(u))^3 := \int_0^1 e^{-|x - y|} u(y)dy + (\arctan(u))^3.
$$

Since the function $u \to \arctan^3 u$ is increasing on $\mathbb{R}$, one has

$$
\langle (\arctan(u))^3 - (\arctan(v))^3, u - v \rangle \geq 0, \quad \forall u, v \in H.
$$

Moreover,

$$
e^{-|x|} = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{i\lambda x} \frac{e^{i\lambda x}}{1 + \lambda^2} d\lambda.
$$

Therefore, $\langle B(u - v), u - v \rangle \geq 0$, so

$$
\langle F(u - v), u - v \rangle \geq 0, \quad \forall u, v \in H.
$$

Thus, $F$ is a monotone operator. Note that

$$
\langle (\arctan(u))^3 - (\arctan(v))^3, u - v \rangle = 0 \iff u = v \quad a.e.
$$
Therefore, the operator $F$, defined in (1.1), is injective and equation (1.1), with this $F$, has at most one solution.

The Fréchet derivative of $F$ is

$$F'(u)w = \frac{3\left(\arctan(u)\right)^2}{1+u^2}w + \int_0^1 e^{-|x-y|}w(y)dy.$$  

If $u(x)$ vanishes on a set of positive Lebesgue measure, then $F'(u)$ is not boundedly invertible. If $u \in C[0,1]$ vanishes even at one point $x_0$, then $F'(u)$ is not boundedly invertible in $H$.

In numerical implementation of the DSM, one often discretizes the Cauchy problem (3.1) and gets a system of ordinary differential equations (ODEs). Then, one can use numerical methods for solving ODEs to solve the system of ordinary differential equations obtained from discretization. There are many numerical methods for solving ODEs (see, e.g., [2]).

In practice one does not have to compute $u_\delta(t_\delta)$ exactly but can use an approximation to $u_\delta(t_\delta)$ as a stable solution to equation (1.1). To calculate such an approximation, one can use, for example, the iterative scheme

$$u_{n+1} = u_n - (F'(u_n) + a_n I)^{-1}(F(u_n) + a_n u_n - f_\delta),$$

$$u_0 = 0,$$

and stop iterations at $n := n_\delta$ such that the following inequality holds:

$$\|F(u_{n_\delta}) - f_\delta\| < C\delta^\gamma, \quad \|F(u_n) - f_\delta\| \geq C\delta^\gamma, \quad n < n_\delta, \quad C > 1, \quad \gamma \in (0,1).$$

The existence of the stopping time $n_\delta$ is proved in [3, p. 733] and the choice $u_0 = 0$ is also justified in this paper. Iterative scheme (4.6) and stopping rule (4.7) are used in the numerical experiments. We proved in [3, p. 733] that $u_{n_\delta}$ converges to $u^*$, a solution of (1.1). Since $F$ is injective as discussed above, we conclude that $u_{n_\delta}$ converges to the unique solution of equation (1.1) as $\delta$ tends to 0. The accuracy and stability are the key issues in solving the Cauchy problem. The iterative scheme (4.6) can be considered formally as the explicit Euler’s method with the stepsize $h = 1$ (see, e.g., [2]). There might be other iterative schemes which are more efficient than scheme (4.6), but this scheme is simple and easy to implement.

Integrals of the form $\int_0^1 e^{-|x-y|}h(y)dy$ in (4.1) and (4.5) are computed by using the trapezoidal rule. The noisy function used in the test is

$$f_\delta(x) = f(x) + \kappa f_{\text{noise}}(x), \quad \kappa > 0.$$ 

The noise level $\delta$ and the relative noise level are defined by the formulas

$$\delta = \kappa \|f_{\text{noise}}\|, \quad \delta_{\text{rel}} := \frac{\delta}{\|f\|}.$$ 

In the test $\kappa$ is computed in such a way that the relative noise level $\delta_{\text{rel}}$ equals some desired value, i.e.,

$$\kappa = \frac{\delta}{\|f_{\text{noise}}\|} = \frac{\delta_{\text{rel}} \|f\|}{\|f_{\text{noise}}\|}.$$ 

We have used the relative noise level as an input parameter in the test.

In all the figures the $x$-variable runs through the interval $[0,1]$, and the graphs represent the numerical solutions $u_{\text{DSM}}(x)$ and the exact solution $u_{\text{exact}}(x).$
In the test we took \( h = 1, \) \( C = 1.01, \) and \( \gamma = 0.99. \) The exact solution in the test is

\[
 u_e(x) = \begin{cases} 
 0 & \text{if } \frac{1}{3} \leq x \leq \frac{2}{3}, \\
 1 & \text{otherwise},
\end{cases}
\]

here \( x \in [0, 1], \) and the right-hand side is \( f = F(u_e). \) As mentioned above, \( F'(u) \) is not boundedly invertible in any neighborhood of \( u_e. \)

It is proved in [3] that one can take \( a_n = \frac{d}{1 + n}, \) and \( d \) is sufficiently large. However, in practice, if we choose \( d \) too large, then the method will use too many iterations before reaching the stopping time \( n_\delta \) in (4.7). This means that the computation time will be large in this case. Since

\[
 \|F(V_\delta) - f_\delta\| = O(a(t_\delta))\|V_\delta\|,
\]

and \( \|V_\delta(t_\delta) - u_\delta(t_\delta)\| = O(a(t_\delta))\|V_\delta\|, \) we have

\[
 C\delta^\gamma = \|F(u_\delta(t_\delta)) - f_\delta\| \leq a(t_\delta)\|V_\delta\| + O(a(t_\delta)),
\]

and we choose

\[
 d = C_0\delta^\gamma, \quad C_0 > 0.
\]

In the experiments our method works well with \( C_0 \in [7, 10]. \) In numerical experiments, we found out that the method diverged for smaller \( C_0. \) In the test we chose \( a_n \) by the formula \( a_n := C_0 \frac{99}{n^2}. \) The number of nodal points, used in computing integrals in (4.1) and (4.5), was \( N = 100. \) The accuracy of the solutions obtained in the tests with \( N = 30 \) and \( N = 50 \) was slightly less accurate than the one for \( N = 100. \)

Numerical results for various values of \( \delta_{rel} \) are presented in Table 1. In this experiment, the noise function \( f_{\text{noise}} \) is a vector with random entries normally distributed, with mean value 0 and variance 1. Table 1 shows that the iterative scheme yields good numerical results.

**Table 1.** Results when \( C_0 = 7, \) \( N = 100 \) and \( u = u_e. \)

<table>
<thead>
<tr>
<th>( \delta_{rel} )</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.003</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>57</td>
<td>57</td>
<td>58</td>
<td>58</td>
<td>59</td>
</tr>
<tr>
<td>( |u_{DSM} - u_{exact}|/|u_{exact}| )</td>
<td>0.1437</td>
<td>0.1217</td>
<td>0.0829</td>
<td>0.0746</td>
<td>0.0544</td>
</tr>
</tbody>
</table>

Figure 1 presents the numerical results when \( N = 100 \) and \( C_0 = 7 \) with \( \delta_{rel} = 0.01 \) and \( \delta_{rel} = 0.005. \) The numbers of iterations for \( \delta = 0.01 \) and \( \delta = 0.005 \) were 57 and 58, respectively.

Figure 2 presents the numerical results when \( N = 100 \) and \( C_0 = 7 \) with \( \delta = 0.003 \) and \( \delta = 0.001. \) In these cases, it took 58 and 59 iterations to get the numerical solutions for \( \delta_{rel} = 0.003 \) and \( \delta_{rel} = 0.001, \) respectively.

We also carried out numerical experiments with \( u(x) \equiv 1, \) \( x \in [0, 1], \) as the exact solution. Note that \( F'(u) \) is boundedly invertible at this exact solution. However, in any arbitrarily small (in \( L^2 \) norm) neighborhood of this solution, there are infinitely many elements \( u \) at which \( F'(u) \) is not boundedly invertible, because, as we have pointed out earlier, \( F'(u) \) is not boundedly invertible if \( u(x) \) is continuous and vanishes at some point \( x \in [0, 1]. \) In this case one cannot use the usual methods like Newton’s method or the Newton-Kantorovich method. Numerical results for this experiment are presented in Table 2.
Figure 1. Plots solutions obtained by the DSM when $N = 100$, $\delta_{rel} = 0.01$ (left) and $\delta_{rel} = 0.005$ (right).

Figure 2. Plots solutions obtained by the DSM when $N = 100$, $\delta_{rel} = 0.003$ (left) and $\delta_{rel} = 0.001$ (right).

Table 2. Results when $C_0 = 4$, $N = 50$ and $u(x) \equiv 1$, $x \in [0, 1]$.

<table>
<thead>
<tr>
<th>$\delta_{rel}$</th>
<th>0.05</th>
<th>0.03</th>
<th>0.02</th>
<th>0.01</th>
<th>0.003</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>28</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>$\frac{|u_{DSM} - u_{exact}|}{|u_{exact}|}$</td>
<td>0.0770</td>
<td>0.0411</td>
<td>0.0314</td>
<td>0.0146</td>
<td>0.0046</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

From Table 2 one concludes that the method works well in this experiment.
4.2. An experiment with an operator defined on a dense subset of $H = L^2[0,1]$. Our second numerical experiment with the equation $F(u) = f$ deals with the operator $F$ which is not defined on all of $H = L^2[0,1]$, but on a dense subset $D = C[0,1]$ of $H$:

$$F(u) := B(u) + u^3 := \int_{0}^{1} e^{-|x-y|} u(y)dy + u^3.$$  

(4.9)

Therefore, the assumptions of Theorem 3.1 are not satisfied. Our goal is to show by this numerical example, that numerically our method may work for an even wider class of problems than that covered by Theorem 3.1.

The operator $B$ is compact in $H = L^2[0,1]$. The operator $u \mapsto u^3$ is defined on a dense subset $D$ of $L^2[0,1]$, for example, on $D := C[0,1]$. If $u, v \in D$, then

$$\langle u^3 - v^3, u - v \rangle = \int_{0}^{1} (u^3 - v^3)(u - v)dx \geq 0.$$  

(4.10)

This and the inequality $\langle B(u - v), u - v \rangle \geq 0$, followed from equality (4.3), imply

$$\langle F(u - v), u - v \rangle \geq 0, \quad \forall u, v \in D.$$  

Note that the equal sign of inequality (4.10) happens iff $u = v$ a.e. in Lebesgue measure. Thus, $F$ is injective. Therefore, the element $u_n \in D$ obtained from the iterative scheme (4.6) and the stopping rule (4.7) converges to the exact solution $u_e$ as $\delta$ goes to 0.

Note that $D$ does not contain subsets open in $H = L^2[0,1]$, i.e., it does not contain interior points of $H$. This is a reflection of the fact that the operator $G(u) = u^3$ is unbounded on any open subset of $H$. For example, in any ball $\|u\| \leq C$, $C = const > 0$, where $\|u\| := \|u\|_{L^2[0,1]}$, there is an element $u$ such that $\|u^3\| = \infty$. As such an element one can take, for example, $u(x) = c_1 x^{-b}$, $\frac{1}{2} < b < \frac{1}{2}$. Here $c_1 > 0$ is a constant chosen so that $\|u\| \leq C$. The operator $u \mapsto F(u) = G(u) + B(u)$ is maximal monotone on $D_F := \{u : u \in H, F(u) \in H\}$ (see [1] p. 102), so that equation (2.11) is uniquely solvable for any $f_b \in H$.

The Fréchet derivative of $F$ is

$$F'(u)w = 3u^2 w + \int_{0}^{1} e^{-|x-y|} w(y)dy.$$  

(4.11)

If $u(x)$ vanishes on a set of positive Lebesgue measure, then $F'(u)$ is obviously not boundedly invertible. If $u \in C[0,1]$ vanishes even at one point $x_0$, then $F'(u)$ is not boundedly invertible in $H$.

We also use the iterative scheme (4.6) with the stopping rule (4.7).

We use the same exact solution $u_e$ as in (4.8). The right-hand side $f$ is computed by $f = F(u_e)$. Note that $F'$ is not boundedly invertible in any neighborhood of $u_e$.

In experiments we found that our method works well with $C_0 \in [1, 4]$. Indeed, in the test we chose $a_n$ by the formula $a_n := C_0 \frac{N^2}{N}$. The number of node points used in computing integrals in (4.11) and (4.2) was $N = 30$. In the test, the accuracy of the solutions obtained when $N = 30$, $N = 50$ were slightly less accurate than the one when $N = 100$.

Numerical results for various values of $\delta_{rel}$ are presented in Table 3. In this experiment, the noise function $f_{\text{noise}}$ is a vector with random entries normally distributed of mean 0 and variance 1. Table 3 shows that the iterative scheme yields good numerical results.
Table 3. Results when $C_0 = 2$ and $N = 100$.

<table>
<thead>
<tr>
<th>$\delta_{rel}$</th>
<th>0.02</th>
<th>0.01</th>
<th>0.005</th>
<th>0.003</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>$|u_{DSM} - u_{exact}|$</td>
<td>0.1387</td>
<td>0.1281</td>
<td>0.0966</td>
<td>0.0784</td>
<td>0.0626</td>
</tr>
</tbody>
</table>

Figure 3 presents the numerical results when $f_{noise}(x) = \sin(3\pi x)$ for $\delta_{rel} = 0.02$ and $\delta_{rel} = 0.01$. The number of iterations when $C_0 = 2$ for $\delta_{rel} = 0.02$ and $\delta_{rel} = 0.01$ were 16 and 17, respectively.

![Figure 3](image)

Figure 4 presents the numerical results when $f_{noise}(x) = \sin(3\pi x)$ with $\delta_{rel} = 0.003$ and $\delta_{rel} = 0.001$. We also used $C_0 = 2$. In these cases, it took 17 and 18 iterations to give the numerical solutions for $\delta_{rel} = 0.003$ and $\delta_{rel} = 0.001$, respectively.

We have included the results of the numerical experiments with $u(x) \equiv 1, x \in [0,1]$, as the exact solution. The operator $F'(u)$ is boundedly invertible in $L^2([0,1])$ at this exact solution. However, in any arbitrarily small $L^2$-neighborhood of this solution, there are infinitely many elements $u$ at which $F'(u)$ is not boundedly invertible as was mentioned above. Therefore, even in this case one cannot use the usual methods such as Newton’s method or the Newton-Kantorovich method. Numerical results for this experiment are presented in Table 4.

Table 4. Results when $C_0 = 1$, $N = 30$ and $u(x) = 1, x \in [0,1]$.

<table>
<thead>
<tr>
<th>$\delta_{rel}$</th>
<th>0.05</th>
<th>0.03</th>
<th>0.02</th>
<th>0.01</th>
<th>0.003</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$|u_{DSM} - u_{exact}|$</td>
<td>0.0436</td>
<td>0.0245</td>
<td>0.0172</td>
<td>0.0092</td>
<td>0.0026</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Figure 4. Plots solutions obtained by the DSM with $f_{\text{noise}}(x) = \sin(3\pi x)$ when $N = 100$, $\delta_{rel} = 0.003$ (left) and $\delta_{rel} = 0.001$ (right).

From the numerical experiments we can conclude that the method works well in this experiment. Note that the function $F$ used in this experiment is not defined on the whole space $H = L^2[0, 1]$ but defined on a dense subset $D = C[0, 1]$ of $H$.

Added after proof

While this paper was under consideration the following papers on the DSM have appeared:


References


Mathematics Department, Kansas State University, Manhattan, Kansas 66506-2602
E-mail address: nguyenhs@math.ksu.edu

Mathematics Department, Kansas State University, Manhattan, Kansas 66506-2602
E-mail address: ramm@math.ksu.edu