THE COMPUTATION OF CONVEX INVARIANT SETS VIA NEWTON’S METHOD

R. Baier
Chair of Applied Mathematics
University of Bayreuth
95440 Bayreuth, Germany

M. Dellnitz, M. Hessel-von Molo and S. Sertl
Chair of Applied Mathematics
University of Paderborn
33098 Paderborn, Germany

I. G. Kevrekidis
Department of Chemical and Biological Engineering
Princeton University
Princeton, NJ 08544, USA

(Communicated by Bernd Krauskopf)

ABSTRACT. In this paper we present a novel approach to the computation of convex invariant sets of dynamical systems. Employing a Banach space formalism to describe differences of convex compact subsets of \( \mathbb{R}^n \) by directed sets, we are able to formulate the property of a convex, compact set to be invariant as a zero-finding problem in this Banach space. We need either the additional restrictive assumption that the image of sets from a subclass of convex compact sets under the dynamics remains convex, or we have to convexify these images. In both cases we can apply Newton’s method in Banach spaces to approximate such invariant sets if an appropriate smoothness of a set-valued map holds. The theoretical foundations for realizing this approach are analyzed, and it is illustrated first by analytical and then by numerical examples.

1. Introduction. One of the most important subjects in the numerical treatment of dynamical systems is the efficient and reliable computation of invariant sets. Therefore this topic has already been studied extensively. For the computation of (time independent) steady state solutions — or fixed points in the case of discrete time — one can use standard contraction mapping techniques like (quasi-) Newton methods. But already the numerical computation of invariant sets such as invariant manifolds is much more challenging, and several techniques have been suggested in the last two decades of the 20th century — see e.g. [9, 20] or [29] for a survey. In that period also the first articles were published where systems with even more
complicated dynamics have been of interest. In fact, in [25, 15] numerical methods have been proposed for the computation of invariant tori, and in [12] a set oriented approach has been suggested for the computation of general global attractors that was extended in [13].

Up to now this set oriented approach for the computation of invariant sets is essentially based on “fixed set” iterations (see [34, 17] and references therein) or on cell mapping techniques as developed e.g. in [21, 32]. Set iterations in Euler-type form and invariant sets under one-sided Lipschitz conditions are studied in [17], i.e. the iteration for a differential inclusion with set-valued right-hand side \( F(\cdot) \) is as follows:

\[
X_{k+1} = G_h(X_k) \quad \text{with} \quad G_h(x) = \{ x + h\xi : \xi \in F(x) \}, \quad G_h(X_k) = \bigcup_{x_k \in X_k} \{ G_h(x_k) \}.
\]

In [12] a particularly efficient numerical realization of cell mapping techniques has been presented. This algorithm has been implemented in the software package GAIO\(^1\)[11]. It combines the cell mapping approach with a subdivision-selection-technique for the refinement of a box collection. Starting with a coarse approximation of the invariant set to be computed, it iteratively refines the boxes in the current collection by subdivision, computes their images, selects suitable boxes for further refinement and discards the others. Therefore, this technique can be viewed as a numerical realization of a fixed point iteration for invariant sets. Viewed in this light, this approach has two particular drawbacks:

(a) It only allows to approximate invariant sets which are stable “fixed sets” for this fixed point iteration, and
(b) the speed of convergence is only linear and depends on the dynamical attractiveness of the underlying invariant set.

In this article we propose a new set oriented approach for the computation of convex invariant sets which, in principle, overcomes these two drawbacks if the images of the set iterates remain convex or at least lie in the closed linear span of (convex) compatible sets. The underlying idea can be illustrated as follows: Suppose that one would like to compute a fixed point of a differentiable map \( g \), that is, a point \( p \in \mathbb{R}^n \) such that \( g(p) = p \). In this case one does not have to rely on a fixed point iteration. Rather one can apply Newton’s method to the equation \( g(x) - x = 0 \) in order to compute these points. Observe that for this approach the stability property of \( p \) does neither affect the computability of \( p \) nor the speed of convergence of Newton’s method. Our idea is to proceed in a similar way for the case of convex invariant sets — i.e. convex sets \( A \subset \mathbb{R}^n \) for which \( g(A) = A \). Thus, we would like to find a solution of

\[
g(A) - A = \{ 0 \}
\]

via a set-valued Newton’s method. In analogy to the case of fixed points this would allow us to compute unstable or hyperbolic invariant sets and the speed of convergence would be superior to fixed point iterations as implemented in GAIO. However, a priori it is not obvious at all how to interpret the expression, in particular the minus sign, in (1), and which set-valued derivative to apply to make use of Newton’s method in this context.

Let us note that the proposed set-valued Newton’s method differs from the generalized Newton-type methods applied to nonsmooth equations (see e.g. [28] for many

\(^1\)GAIO can be obtained from its authors upon request to gaio@djs2.de.
references) in which subdifferentials or generalized gradients appear and cause a set-valuedness.

Set-valued versions of Newton’s method are mainly studied for intervals only. Classical implementations as e.g. in [30, 31, 19] use arithmetic operations which are adequate for containing all possible results (motivated by rounding errors in computers) but are not compatible with a vector space approach. An exception are the works [33, 16], where operations for directed intervals (one-dimensional directed sets) are used in the Newton’s method.

In two pioneering works, a set-valued Newton’s method in higher dimensions is studied: the first one introduces the concept for a set-valued Newton’s method applied to mutational equations in [1, Sec. 2.5], the second one in [42] studies the problem $0 \in F(x)$ and applies the viability approach to a suitable differential inclusion together with homotopic path algorithms.

The essential problem to define a set-valued Newton’s method closer to the pointwise case is that conventional notions for differences of sets do not suffice for formulating (1). While the Minkowski addition of sets

$$A + B = \{a + b \mid a \in A, b \in B\}$$

for $A, B \subset \mathbb{R}^n$ and the multiplication with non-negative scalars

$$s \cdot A = \{sa \mid a \in A\}$$

is well-accepted in set-valued numerical analysis, the multiplication with a negative scalar $s \in \mathbb{R}$ is problematic, since known laws from $\mathbb{R}^n$ cannot be extended to it. It was remarked in [7] that usual convex differences like the geometric (Hadwiger-Pontryagin) difference “−∗” in [18, 39] satisfy only inclusions for the associative law and are in general either too small, i.e.

$$(A \oplus B) + B \subseteq A,$$

or too big as the Demyanov difference “−·” in [41, 14], since

$$(A - \cdot B) + B \supseteq A.$$ The simplest approach with the algebraic difference “⊖” in

$$A \ominus B := \{a - b : a \in A, b \in B\}$$

yields a superset of the Demyanov difference and can only guarantee

$$B \ominus B \supseteq \{0\}$$

in general. These examples indicate that these three differences cannot be inverse to the Minkowski sum in general. For all sets except singletons, the scalar multiplication “⊙” of $B$ by the negative value $-1$ defined as $\{-b : b \in B\}$ will not create an inverse as well, since

$$B + (-1) \odot B = B \ominus B \supseteq \{0\}. $$

Obviously, the notion of a difference is essential for the definition of a derivative of a set-valued map. Therefore, it seems understandable that most of the derivative notions have limited calculus rules compared to the point-wise case, often with inclusions replacing equalities.

An exception is given by the quasidifferential which uses the embedding into the Banach space of pairs of convex compact sets (see [14, 35]), but no visualization of these pairs of sets as nonconvex subsets of $\mathbb{R}^n$ is applied there. In applications (see [14]), either the geometric difference is used (e.g. in optimality conditions for
unconstrained minimization problems) or the Demyanov difference (e.g. in estimating or obtaining Clarke’s generalized gradients from the quasidifferential of a nonsmooth function).

As in the case of subdifferentials for DC (difference of convex) functions (see [4]), there is a need for nonconvex differences or derivatives of convex sets resp. set-valued maps with convex images. Set-valued derivatives based on directed sets can be found in [3] for the study of several first-order approximations of set-valued maps and in [36, 6] for set-valued polynomial interpolation of higher order and for set-valued Hermite interpolation. In both articles, set-valued divided differences are also defined and the connection to set-valued derivatives is illustrated. Essential ideas and results from this article were used in [5], where divided differences for multivariate set-valued maps are applied to the secant method.

It is the purpose of this article to set up a mathematical framework for a set-valued Newton’s method and to develop a corresponding numerical scheme. The Banach space of directed sets [7, 8] provides an extension of the arithmetical operations of sets (addition, multiplication with non-negative scalars) and introduces a new interpretation of the multiplication with negative scalars. This leads to a notion of difference that is inverse with respect to the (extended) Minkowski sum:

\[(J_n(A) - J_n(B)) + J_n(B) = J_n(A),\]

\[J_n(B) + (-1) \cdot J_n(B) = J_n(\{0\}) = 0.\]

Here, \(J_n\) denotes the embedding of convex, compact subsets of \(\mathbb{R}^n\) into the Banach space of directed sets. Furthermore, directed sets can be visualized as usually nonconvex subsets of \(\mathbb{R}^n\) with three parts, the positive part (with outer normals), the negative part (with inner normals) and the (nonconvex) mixed-type part.

The developments in this publication can be viewed in a broader context. In some applications, the map \(g\) for which invariant sets are to be computed may e. g. be implemented by a “timestepper”, i. e. a numerical algorithm that integrates a system of ODEs or a PDE over one time step (see e. g. [45]). In many specialized application contexts, simulation algorithms based on timesteppers have been developed, fine-tuned and validated in recent years. In multiscale modeling of complex systems, this timestepper based approach has been the basis of computational algorithms that circumvent the derivation of explicit coarse-grained equations; instead these algorithms perform advanced coarse-grained numerical tasks (such as e. g. the computation of fixed points or bifurcations) acting on a microscopic/atomistic/fine scale simulator directly. This idea forms the basis of the so-called Equation-Free framework [44, 27, 26, 43] for complex/multiscale systems modeling. In this light, the Newton algorithm we propose here may just be the forerunner of a wider variety of computer-assisted tasks that use set-valued representations of a map (legacy timestepper) in similar ways as in this article. One of these may be projective integration, where the time derivative required in an initial value problem solver is estimated “on the fly” from the differences between the results of short simulation “bursts”.

The article is structured as follows. In Section 2 we briefly introduce the mathematical background material. In particular, in Section 2.1 we describe the construction of certain Banach spaces, the elements of which are generalizations of convex compact sets called directed sets. In Section 2.2 we briefly recall some results on Newton iterations in Banach spaces that will be needed later. In Section 3, we analyze the differentiation of maps of directed sets. This will be the main prerequisite
for the application of Newton’s method to such maps in Section 4. In Section 5 we describe a particular numerical realization of this approach yielding a novel numerical method for the computation of convex, compact invariant sets. Finally we demonstrate the applicability and usefulness of this method both by numerical examples in Section 6.

2. Preliminaries.

2.1. Banach spaces of directed sets. In this section we briefly review the theoretical results which will be necessary for our considerations. Here we essentially summarize parts of the developments in [7, 8], where a Banach space calculus for convex, compact sets — called the space of directed sets — is described.

It is our aim to describe a method to compute convex, compact invariant sets of dynamical systems that directly operates on (suitable representations of) subsets of $\mathbb{R}^n$, cutting short the approximation of such a set by a large (finite) collection of its elements. To this end, it is necessary to have a vector space structure on a set of subsets of $\mathbb{R}^n$ that extends the known vector space structure on $\mathbb{R}^n$. Such a structure is not canonically given — the natural extensions of the vector space structure on $\mathbb{R}^n$ to sets of subsets of $\mathbb{R}^n$ given by the set operations (2)–(3) have properties that do not suffice for our purposes. In particular, there is no distributive law, and typically there are no additive inverses. For example, if $B = B_1(0) \subset \mathbb{R}^n$ is the Euclidean unit ball, then one has $B = (-1) \odot B$ and thus

\[
\{0\} = 0 \cdot B = (1 - 1) \cdot B \neq B + (-1) \odot B = B + B = 2 \cdot B.
\]

In [7] a Banach space of functions is constructed, some of whose elements can be interpreted as representing sets in $\mathbb{R}^n$, in such a way that the Banach space arithmetic emulates the basic arithmetic on sets given above. The price one has to pay for this is that one has to restrict to convex, compact sets in $\mathbb{R}^n$. Indeed, the key to this construction is the fact that convex, compact sets are already completely described by their boundary. It uses a careful parametrization of the boundary of a convex set which also takes into account a certain sense of direction which allows to distinguish “positively oriented” sets from “negatively oriented” counterparts as well as nonconvex cases with an appearing mixed-type part (boundary points that lie both in half-spaces with positive and negative “orientations”) in such a way that the sum of these two is zero, that is, the set containing only the vector $0 \in \mathbb{R}^n$.

In the following, let $\mathcal{C}(\mathbb{R}^n)$ denote the set of convex, compact subsets of $\mathbb{R}^n$. For the definition of the space of directed sets we will need the notions of the support function and the supporting faces of a convex set.

Definition 2.1. Let $A \in \mathcal{C}(\mathbb{R}^n)$. The function

\[
\delta^* (\cdot, A) : S^{n-1} \to \mathbb{R}, \ l \mapsto \max_{a \in A} (a, l)
\]

defined on the unit sphere $S^{n-1}$ in $\mathbb{R}^n$ is called the support function of $A$. The map

\[
Y (\cdot, A) : S^{n-1} \to \mathcal{C}(\mathbb{R}^n), \ l \mapsto \{a \in A \mid (a, l) = \delta^* (l, A)\}
\]

defines the supporting face of $A$ in direction $l$.

The supporting face $Y (l, A)$ is a convex, compact set that is orthogonal to $l$ and is contained in the boundary of $A$. It can be described geometrically as the intersection of $A$ with the hyperplane orthogonal to $l$ that intersects the boundary, but not the interior of $A$. Correspondingly, it can be identified with a convex,
compact set in $\mathbb{R}^{n-1}$. This identification is the first main building block in the definition of directed sets. The second one is the fact that one can reconstruct a convex set from its support function. By construction, we have for $A \in C(\mathbb{R}^n)$ that $A = \bigcap_{t \in S^{n-1}} \{x \mid \langle x, l \rangle \leq \delta^*(l, A)\}$, see e.g. [40, Theorem 13.1]. (Here and in the following, $S^{n-1}$ denotes the unit sphere in $\mathbb{R}^n$.)

**Definition 2.2.** The space of pre-directed intervals is

$$\mathcal{D}(\mathbb{R}) = \{\tilde{A}_1 \mid \tilde{A}_1 : S^0 \to \mathbb{R}\}$$

with the norm $\|\tilde{A}_1\|_1 = \sup_{t \in S^0} |\tilde{A}_1(t)|$. For $n > 1$, the space of $n$-dimensional pre-directed sets is

$$\mathcal{D}(\mathbb{R}^n) = \left\{ (\tilde{A}_{n-1}, a_n) \mid a_n \in C^0(S^{n-1}, \mathbb{R}), \right.\left. \tilde{A}_{n-1} : S^{n-1} \to \mathcal{D}(\mathbb{R}^{n-1}) \right\}$$

with the norm $\|(\tilde{A}_{n-1}, a_n)\|_n = \sup_{t \in S^n} \max\{\|a_n(t)\|, \|\tilde{A}_{n-1}(t)\|_{n-1}\}$.

As a space of functions on $S^{n-1}$ with values in a linear space, $\mathcal{D}(\mathbb{R}^n)$ is a normed linear space with the usual pointwise definitions of addition and scalar multiplication. It was shown in [7] that $(\mathcal{D}(\mathbb{R}^n), \| \cdot \|_n)$ is a Banach space. In the following, we will drop the index $n$ in the norm.

Briefly, the second component of a pre-directed set can be thought of as describing the support function of the underlying convex compact set, and the first component as describing the supporting faces. The following definition of a mapping that embeds $C(\mathbb{R}^n)$ into $\mathcal{D}(\mathbb{R}^n)$ will make this more precise.

**Definition 2.3.** The embedding $J_n : C(\mathbb{R}^n) \to \mathcal{D}(\mathbb{R}^n)$ of the convex compact sets into $\mathcal{D}(\mathbb{R}^n)$ is defined as follows:

(i) For $n = 1$ the embedding $J_1 : C(\mathbb{R}) \to \mathcal{D}(\mathbb{R})$ is defined by

$$J_1([a, b]) = (\delta^*(l, [a, b]))_{l=\pm1} = (-a, b).$$

(ii) For $n \geq 2$ the embedding $J_n : C(\mathbb{R}^n) \to \mathcal{D}(\mathbb{R}^n)$ is defined by

$$A \mapsto (\tilde{A}_{n-1}, a_n),$$

where $a_n(l) = \delta^*(l, A)$ is the support function of $A$ and

$$\tilde{A}_{n-1}(l) = J_{n-1}(\pi_{n-1}(R_n^l(Y(l, A) - \delta^*(l, A) \cdot l))),$$

where $\pi_{n-1} : \mathbb{R}^n \to \mathbb{R}^{n-1}$ is the natural projection given by

$$\pi_{n-1}(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1})^T$$

and $R_n^l \in SO(n)$ is an arbitrary but fixed rotation on $\mathbb{R}^n$ that maps $l$ onto the unit vector $(0, \ldots, 0, 1)^T$.

The right hand side of (4) is best understood by considering the elements from the innermost terms outward. The set $Y(l, A) - \delta^*(l, A) \cdot l$ is the result of translating the supporting face of $A$ in direction $l$ along that direction into the orthogonal complement of $l$ in $\mathbb{R}^n$. The application of $R_n^l$ rotates this subspace into the subspace spanned by the first $n-1$ unit vectors, and the subsequent projection just “throws away” the $n$-th component of the resulting vectors. The resulting set is thus the supporting face $Y(l, A)$ viewed as a convex, compact set in $\mathbb{R}^{n-1}$, which by application of $J_{n-1}$ is finally embedded into $\mathcal{D}(\mathbb{R}^{n-1})$. 
Definition 2.4. We denote the range of \( J_n \) by \( \overline{\mathcal{C}}(\mathbb{R}^n) = J_n(\mathcal{C}(\mathbb{R}^n)) \subset \overline{\mathcal{D}}(\mathbb{R}^n) \). Furthermore, we define the space of directed sets for dimension \( n = 1 \) as \( \mathcal{D}(\mathbb{R}) = \overline{\mathcal{D}}(\mathbb{R}) \) and for higher dimensions as the closure of the linear span \( \langle \ldots \rangle \) of \( \overline{\mathcal{C}}(\mathbb{R}^n) \) with respect to the norm on \( \overline{\mathcal{D}}(\mathbb{R}^n) \):

\[
\mathcal{D}(\mathbb{R}^n) := \overline{\langle \mathcal{C}(\mathbb{R}^n) \rangle}.
\]

As a closed linear subspace of a Banach space, \( \mathcal{D}(\mathbb{R}^n) \) is also a Banach space. It is straightforward to check that \( J_n \) has the following useful property:

Lemma 2.5. Let \( A, B \in \mathcal{C}(\mathbb{R}^n) \) and \( \alpha > 0 \). Then

\[
J_n(A + \alpha B) = J_n(A) + \alpha J_n(B),
\]

where the operations on subsets of \( \mathbb{R}^n \) are defined as in (2) and (3).

This means in particular that \( \overline{\mathcal{C}}(\mathbb{R}^n) \) is a cone in the Banach space of directed sets. The embedding \( J_n(\cdot) \) has a counterpart that maps elements of \( \mathcal{D}(\mathbb{R}^n) \) to subsets of \( \mathbb{R}^n \). This “visualization” mapping is defined in the following

Definition 2.6. Let \( \overrightarrow{A} = (\overrightarrow{A_{n-1}}, a_n) \in \mathcal{D}(\mathbb{R}^n) \). The visualization of \( \overrightarrow{A} \) is the set \( V_n(\overrightarrow{A}) = P_n(\overrightarrow{A}) \cup N_n(\overrightarrow{A}) \cup M_n(\overrightarrow{A}) \), where the positive (proper) part \( P_n(\overrightarrow{A}) \), the negative (improper) part \( N_n(\overrightarrow{A}) \) and the mixed-type part \( M_n(\overrightarrow{A}) \) are defined as follows:

\[
P_n(\overrightarrow{A}) = \bigcap_{l \in S} \{ x \in \mathbb{R}^n \mid \langle l, x \rangle \leq a_n(l) \},
\]

\[
N_n(\overrightarrow{A}) = \ominus \bigcap_{l \in S} \{ x \in \mathbb{R}^n \mid \langle l, x \rangle \leq -a_n(l) \},
\]

where \( \ominus A \) for a set \( A \in \mathbb{R}^n \) is the set \( \ominus A = \{ -x \mid x \in A \} \).

The mixed-type part \( M_n \) is defined via the boundary \( B_n(\overrightarrow{A}) \).

For \( n = 1 \) we set\(^2\) \( M_1(\overrightarrow{A}) = \emptyset \), and for \( n \geq 2 \) the boundary set is formed by all reprojections of the visualizations of the \((n-1)\)-dimensional supporting faces:

\[
B_n(\overrightarrow{A}) = \bigcup_{l \in S} Q_n^{a_n(\cdot)}(V_{n-1}(\overrightarrow{A_{n-1}}(l))),(l),
\]

where \( Q_n^{a_n(\cdot)}(y) = (R_n^{l})^{T} \pi_{n-1}^{a_n(l)}(y) + a_n(l)l \) for \( y \in \mathbb{R}^{n-1} \), with \( \pi_{n-1}^{a_n(l)} \) the left inverse of \( \pi_{n-1}^{a_n(l)} \) that embeds \( \mathbb{R}^{n-1} \) into \( \mathbb{R}^{n} \).

The mixed-type part is now

\[
M_n(\overrightarrow{A}) = B_n(\overrightarrow{A}) \setminus (\partial P_n(\overrightarrow{A}) \cup \partial N_n(\overrightarrow{A})).
\]

Intuitively, the positive part \( P_n(\overrightarrow{A}) \) can be thought of as a convex set with outward normals on its boundary, and the negative part \( N_n(\overrightarrow{A}) \) as a convex set with inward (hence negative) normals. No directed set (with the exception of embedded singleton sets) can have both a non-empty positive and a non-empty negative part [8, Proposition 3.4]. The mixed-type part consists of “remaining boundary points”: points that arise as visualizations of supporting faces, although they do not belong to the boundary of either the positive or the negative part. They can occur e.g. as “residuals” of one set in a difference of their embeddings (see Fig. 1).

\(^2\)This results naturally from using the construction for \( n \geq 2 \) analogously, as with \( B_1(\overrightarrow{A}) = \partial P_1(\overrightarrow{A}) \cup \partial N_1(\overrightarrow{A}) \) one has \( M_1(\overrightarrow{A}) = B_1(\overrightarrow{A}) \setminus (\partial P_1(\overrightarrow{A}) \cup \partial N_1(\overrightarrow{A})) = \emptyset \).
2.2. Newton iterations in Banach spaces. The analysis of Newton iterations in Banach spaces has been the subject of intensive research, most of which goes back to the seminal work by L. V. Kantorovich (see e.g. [22, 23, 24]). Here we briefly review the fundamental results essentially due to Kantorovich, but follow a presentation given by Yamamoto [47].

The basic setting is as follows. Let \( X \) be a Banach space and \( F : U \subseteq X \to X \) be Fréchet-differentiable\(^3\) in an open convex set \( U_0 \subseteq U \). Assume that for some \( x_0 \in U_0 \) (with \( F(x_0) \neq 0 \)) the inverse \( (F'(x_0))^{-1} \) exists as a bounded linear operator on \( X \). Then Newton’s method for finding a solution to

\[
F(x) = 0
\]

starting in \( x_0 \) is defined by the equation

\[
x_{k+1} = x_k - F'(x_k)^{-1}F(x_k),
\]

assuming that \( F'(x_k)^{-1} \) exists for each iterate \( x_k \). For some \( x \in X \) and \( r > 0 \), denote by \( O(x, r) = \{ y \in X \mid \|x - y\| < r \} \) the open ball with radius \( r \) around \( x \). For this setting, one has the following

**Theorem 2.7** (Excerpt from Theorems 2.1 and 3.2 in [47]). In the above described situation, assume that

\[
\|F'(x_0)^{-1}(F'(x) - F'(y))\| \leq K\|x - y\| \quad \forall x, y \in U_0
\]

for some \( K > 0 \). With \( \eta = \|F'(x_0)^{-1}F(x_0)\| \) assume that

\[
h = K\eta \leq \frac{1}{2}.
\]

\(^3\)Recall that if \( X \) and \( Y \) are Banach spaces (the norms of which are both written \( \| \cdot \| \)), and \( U \subseteq X \) is open, then a map \( T : U \to Y \) is called (Fréchet-)differentiable at \( x_0 \in U \) if there is a continuous linear map \( T'(x_0) : X \to Y \) such that

\[
\lim_{h \to 0} \frac{\|T(x_0 + h) - T(x_0) - T'(x_0)h\|}{\|h\|} = 0
\]

holds.
Furthermore, with \( t^* = \frac{2\eta}{1+\sqrt{1-2\eta}} \), let \( O_1 = O(x_1, t^* - \eta) \) and assume that \( \mathcal{O}_1 \subset U_0 \). Then the following statements hold:

(i) The iterates \( x_k \) of (6) are well-defined, \( x_k \in O_1 \) for \( k \geq 1 \), and \( \lim_{k \to \infty} x_k = x* \) exists.

(ii) With \( \tilde{t} = \frac{1+\sqrt{1-2\eta}}{\eta} \), the solution \( x* \) is unique in \( O(x_0, \tilde{t}) \cap U_0 \) if \( 2\eta < 1 \) and in \( \mathcal{O}(x_0, \tilde{t}) \cap U_0 \) if \( 2\eta = 1 \).

(iii) Setting \( d_k = \|x_k - x_{k+1}\| \) and

\[
L_k = \sup_{x,y \in \mathcal{O}_1, x \neq y} \frac{\|F'(x_k)^{-1}(F'(x) - F'(y))\|}{\|x - y\|} \quad \text{for } k \geq 1,
\]

one has

\[
\|x* - x_{k+1}\| \leq \frac{1}{2} L_k \|x* - x_k\|^2
\]

and

\[
\|x* - x_k\| \leq \frac{2d_k}{1 + \sqrt{1 - 2L_k d_k}}.
\]

Thus under the Kantorovich conditions (7) and (8), the Newton iteration is guaranteed to converge quadratically, with errors controllable by \textit{a priori} and \textit{a posteriori} bounds. It is worth noting that (7) is essentially a Lipschitz condition on \( F' \), while (8) can be satisfied by choosing the initial guess \( x_0 \) sufficiently close to the solution \( x* \).

3. Differentiation of maps of directed sets. In order to be able to consider Newton iterations in \( D(\mathbb{R}^n) \), we need notions of differentiability of maps. We use the standard definitions of differentiability for Banach spaces (see e.g. [24]) and consider their meaning in the context of directed sets.

We start with considering the relationship between a linear map on \( \mathbb{R}^n \) and the map on \( D(\mathbb{R}^n) \) generated by it.

**Lemma 3.1.** Let \( M \in \mathbb{R}^{n \times m} \) and \( U, V \in \mathcal{C}(\mathbb{R}^m) \). Then,

\[
\|J_n(MU) - J_n(MV)\| \leq \sqrt{2} \cdot \|M\|_2 \cdot \|J_m(U) - J_m(V)\|
\]

holds with the spectral norm of \( M \), i.e. the l. u. b.-norm \( \|M\|_2 \) with respect to the Euclidean norm.

**Proof.** We use [7, Proposition 4.11] to estimate the norm of the difference of embedded convex sets by their Demyanov distance \( d_D \), which in turn is expressed as the supremum of Hausdorff distances of supporting faces, using [36, Proposition 2.4.5]:

\[
\|J_n(MU) - J_n(MV)\| \leq d_D(MU, MV) = \sup_{l \in S^{n-1}} d_H(Y(l, MU), Y(l, MV))
\]

\[
= \sup_{l \in S^{n-1}} d_H(MY(M^T l, U), MY(M^T l, V))
\]

\[
\leq \|M\|_2 \cdot \sup_{l \in S^{n-1}} d_H(Y(M^T l, U), Y(M^T l, V))
\]

In case \( M^T l = 0 \) we have \( Y(0, U) = U \) and thus

\[
d_H(Y(0, U), Y(0, V)) = d_H(U, V) \leq d_D(U, V).
\]

Otherwise, with \( \zeta = \frac{1}{\|M^T l\|_2} M^T l \in S^{m-1} \) we have \( Y(M^T l, U) = Y(\zeta, U) \) and

\[
d_H(Y(M^T l, U), Y(M^T l, V)) \leq \sup_{\eta \in S^{m-1}} d_H(Y(\eta, U), Y(\eta, V))
\]
so that in both cases
\[ \|J_n(MU) - J_n(MV)\| \leq \|M\|_2 \cdot d_D(U,V) \leq \sqrt{2} \cdot \|M\|_2 \cdot \|J_n(U) - J_n(V)\|. \]

This estimate is used in the following lemma, which will allow us to extend linear maps \(\mathbb{R}^m \to \mathbb{R}^n\) to linear maps \(D(\mathbb{R}^m) \to D(\mathbb{R}^n)\).

**Lemma 3.2.** Let \(M \in \mathbb{R}^{n \times m}\), let \(\overrightarrow{U} \in D(\mathbb{R}^m)\) and for \(k \in \mathbb{N}\) let \(A_k, B_k \in C(\mathbb{R}^m)\) be such that \(\overrightarrow{U} = \lim_{k \to \infty} J_m(A_k) - J_m(B_k)\). Then
\[ \overrightarrow{U} \mapsto M\overrightarrow{U} := \lim_{k \to \infty} J_n(MA_k) - J_n(MB_k) \quad (11) \]
defines a linear map \(M : D(\mathbb{R}^m) \to D(\mathbb{R}^n)\) with the operator norm bounded by \(\sqrt{2} \cdot \|M\|_2\).

**Proof.** Using Lemma 3.1 we see that the right hand side of equation (11) exists and is well-defined. In particular, for any \(r \in \mathbb{N}\) we have
\[ \|J_n(MA_k) - J_n(MB_k)\| = \lim_{k \to \infty} \|J_m(A_k) - J_m(B_k)\| \]
by assumption. Furthermore, if also \(C_k, D_k \in C(\mathbb{R}^m)\) are such that \(\lim_{k \to \infty} J_m(C_k) = J_m(D_k)\), then with
\[ \|J_n(MA_k) - J_n(MB_k)\| = \lim_{k \to \infty} \|J_n(MD_k) + J_m(C_k) - J_m(D_k)\| \]
we see that the left hand side of equation (11) is independent of the choice of the approximating sequence.

To see that \(M : D(\mathbb{R}^m) \to D(\mathbb{R}^n)\) is a linear map, note that equation (11) implies that \(M(-\overrightarrow{U}) = -M(\overrightarrow{U})\) (by changing the roles of \(A_k\) and \(B_k\)), and that for \(\alpha > 0\) Lemma 2.5 implies that \(M(\alpha \overrightarrow{U}) = \alpha M\overrightarrow{U}\). These two facts combined imply \(M(\alpha \overrightarrow{U}) = \alpha M\overrightarrow{U}\) for arbitrary \(\alpha \in \mathbb{R}\). Furthermore, for \(\overrightarrow{V} = \lim_{k \to \infty} J_m(C_k) - J_m(D_k) \in D(\mathbb{R}^n)\) applying Lemma 2.5 to \(M(A_k + C_k)\) and to \(M(B_k + D_k)\) separately allows us to conclude that \(M(\overrightarrow{U} + \overrightarrow{V}) = M\overrightarrow{U} + M\overrightarrow{V}\).

The bound for the operator norm follows directly from Lemma 3.1, i.e.
\[ \|M\overrightarrow{U}\| \leq \sqrt{2} \cdot \|M\|_2 \cdot \|\overrightarrow{U}\|. \]

We now study classes of differentiable maps for simple parametrizations of set-valued maps.

**Lemma 3.3.** Consider \(F : \mathbb{R}^m \to C(\mathbb{R}^n)\), let \(\overrightarrow{F} = J_n \circ F\) be its embedding, and let \(w \in \mathbb{R}^m\).

(i) If \(f : \mathbb{R}^m \to \mathbb{R}^n\) is differentiable and \(F(x) = \{f(x)\}\), then \(\overrightarrow{F}\) is Fréchet-differentiable, and \(\overrightarrow{F}'(x)w = J_n(\{f'(x)w\})\).
Proposition 3.4. \( \text{If for } F_1 : \mathbb{R}^n \rightarrow C(\mathbb{R}^n) \text{ the embedding } \overrightarrow{F}_1(x) = J_n(F_1(x)) \text{ is differentiable and if } r : \mathbb{R}^m \rightarrow \mathbb{R} \text{ is differentiable, then } \overrightarrow{F}(x) = r(x) \cdot J_n(F_1(x)) \text{ is differentiable with } \overrightarrow{F}'(x)w = r'(x) \cdot w \cdot \overrightarrow{F}'_1(x)w. \) \( \text{Especially, if } F_1(x) = U, \text{ then } \overrightarrow{F}'(x)w = r'(x) \cdot w \cdot \overrightarrow{U}. \) \( \text{Furthermore, if } \overrightarrow{F}(x) = r(x) \cdot \overrightarrow{U} \) \( \text{with } \overrightarrow{U} \in D(\mathbb{R}^n), \overrightarrow{F}(\cdot) \text{ is Fréchet-differentiable with the same derivative.} \)

Proof. (i) Here, we use \( J_n(\{v\}) = J_n(\{-v\}) \) \( \text{for a vector } v \in \mathbb{R}^n, \) \( \text{Lemma 2.5 and } [7, \text{Proposition 4.19}] \) \( \text{to obtain} \)

\[
\overrightarrow{F}(y) - \overrightarrow{F}(x) - \overrightarrow{F}'(x)(y - x) = J_n(\{f(y) - f(x) - f'(x)(y - x)\}) = o(\|y - x\|).
\]

(ii) We only prove the formula for \( \overrightarrow{F}(x) = r(x) \cdot J_n(F_1(x)) \):

\[
\begin{align*}
\overrightarrow{F}(y) - \overrightarrow{F}(x) - \overrightarrow{F}'(x)(y - x) &= r(y) \cdot \overrightarrow{F}_1(y) - r(x) \cdot \overrightarrow{F}_1(x) - r'(x) \cdot (y - x) \cdot \overrightarrow{F}_1(x) - r(x) \cdot \overrightarrow{F}'_1(x)(y - x) \\
&= r(y) \cdot (\overrightarrow{F}_1(y) - \overrightarrow{F}_1(x)) + (r(y) - r(x)) \cdot \overrightarrow{F}_1(x) \\
&\quad - r'(x) \cdot (y - x) \cdot \overrightarrow{F}_1(x) - r(x) \cdot \overrightarrow{F}'_1(x)(y - x) \\
&= r(y) \cdot \left( \overrightarrow{F}_1'(x)(y - x) + o(\|y - x\|) \right) - r(x) \cdot \overrightarrow{F}_1'(x)(y - x) \\
&\quad + \left( r'(x) \cdot (y - x) + o(\|y - x\|) \right) \cdot \overrightarrow{F}_1(x) - r'(x) \cdot (y - x) \cdot \overrightarrow{F}_1(x) \\
&= (r(y) - r(x)) \cdot \overrightarrow{F}_1'(x)(y - x) + r(y) \cdot o(\|y - x\|) + o(\|y - x\|) \cdot \overrightarrow{F}_1(x) \\
&= o(\|\overrightarrow{F}_1'(x)\| \cdot \|y - x\|^2) + o(\|y - x\|) = o(\|y - x\|)
\end{align*}
\]

A function with directed intervals as images is differentiable if the boundary functions are both differentiable. First, we study functions that depend on a parameter \( x \in \mathbb{R}^n. \)

**Proposition 3.4.** Let \( \overrightarrow{F} : \mathbb{R}^n \rightarrow D(\mathbb{R}) \) with \( \overrightarrow{F}(x) = \overrightarrow{[f_1(x), f_2(x)]} \) be given with two functions \( f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \) \( i = 1, 2. \) Then, the embedded function \( \overrightarrow{F}(\cdot) \) is differentiable with

\[
\overrightarrow{F}'(x)w = \overrightarrow{[f_1'(x)w, f_2'(x)w]} \quad (w \in \mathbb{R}^n)
\]

if and only if the functions \( f_i(\cdot), \) \( i = 1, 2, \) are differentiable.

Proof. The necessity follows from

\[
\begin{align*}
\overrightarrow{F}(y) - \overrightarrow{F}(x) - \overrightarrow{F}'(x)(y - x) &= [f_1(y) - f_1(x)] - [f_2(y) - f_2(x)] - [f_1'(x)(y - x), f_2'(x)(y - x)] \\
&= [f_1(y) - f_1(x) - f_1'(x)(y - x), f_2(y) - f_2(x) - f_2'(x)(y - x)] \\
&= o(\|y - x\|, o(\|y - x\|) = o(\|y - x\|)
\end{align*}
\]

and Definition 2.2. The linearity of \( w \mapsto \overrightarrow{F}'(x)w \) is a consequence of the arithmetic operations for directed intervals.
Corollary 3.5. Let \( F : X \rightarrow X \) be given as
\[
F(x) = [f_1(x, w), f_2(x, w)]
\]
with two functions \( f_i : X \rightarrow \mathbb{R} \), \( i = 1, 2 \).

Then, the embedded function \( \tilde{F} \) is differentiable with
\[
\tilde{F}(x) = \begin{bmatrix} f_1(x, w) \\ f_2(x, w) \end{bmatrix}
\]
if and only if the functions \( f_i(\cdot), i = 1, 2 \), are differentiable.

The following proposition generalizes Proposition 3.4 \((X = \mathbb{R}^n)\) and Corollary 3.5 \((X = \mathcal{D}(\mathbb{R}))\) to the case \( n \geq 2 \). This lemma is a special case of the following proposition, which for convenience we formulate only for the case \( n \geq 2 \). In fact, the second component of Equation (14) is just a reformulation of Equations (12) and (13) for the case \( n = 1 \).

Proposition 3.6 (Recursive calculation of the Fréchet-derivative). Let \( n \geq 2 \) and let \((X, \| \cdot \|)\) be any Banach space. Consider a map \( F : X \rightarrow \mathcal{D}(\mathbb{R}^n) \) with \( F(x) = (\tilde{F}_{n-1}(x; l), f_n(x; l))_{l \in S^{n-1}} \).

Then, \( F(\cdot) \) is Fréchet-differentiable at \( x \in X \) with
\[
\tilde{F}'(x)w = (\tilde{F}'_{n-1}(x; l)w, f'_n(x; l)w)_{l \in S^{n-1}} \quad (w \in X)
\]
if and only if the map \( \tilde{F}_{n-1}(\cdot; l) \) and the function \( f_n(\cdot; l) \) are differentiable at \( x \) uniformly in \( l \in S^{n-1} \).

Proof. We write \( \mathcal{D} = \tilde{F}'(x)w \in \mathcal{D}(\mathbb{R}^n) \) and consider the two components
\[
\mathcal{D} = (\mathcal{D}_{n-1}(x, w; l), d_n(x, w; l))_{l \in S^{n-1}}.
\]
In order to prove the claim we show that \( \mathcal{D}_{n-1} \) and \( d_n \) are linear first-order approximations of \( \tilde{F}_{n-1} \) and \( f_n \), respectively. Writing \( \mathcal{R} = \frac{1}{\|w\|} \cdot (\tilde{F}(x + w) - \tilde{F}(x) - \tilde{F}'(x)w) \),
we obtain from the definition of the norm on $D(\mathbb{R}^n)$ and from the Fréchet-differentiability of $\overline{F}(\cdot)$ that
\[
\max_{l \in S^{n-1}} \left| \frac{1}{\|y-x\|} (f_n(y;l) - f_n(x;l) - d_n(x,y-x;l)) \right| \leq \|\tilde{R}\| = o(1)
\]
and
\[
\sup_{l \in S^{n-1}} \left| \frac{1}{\|y-x\|} (\overline{F}_{n-1}(y;l) - \overline{F}_{n-1}(x;l) - \overline{D}_{n-1}(x,y-x;l)) \right| \leq \|\tilde{R}\| = o(1).
\]
Hence, it remains to show that $\overline{D}_{n-1}(x, \cdot ; l)$ and $d_n(x, \cdot ; l)$ are linear with respect to the second argument for all $l \in S^{n-1}$. However, this is inherited directly from the linearity of $\overline{F}'(x)$, as for all $\alpha \in \mathbb{R}$, $v, w \in X$ and $l \in S^{n-1}$ we have
\[
\overline{D}_{n-1}(x,v + \alpha w;l) = \overline{D}'(x;l)(v + \alpha w) = \overline{D}'(x;l)v + \alpha \overline{D}'(x;l)w
\]
and hence,
\[
\overline{D}_{n-1}(x,v;l) + \alpha \overline{D}_{n-1}(x,w;l)\]
which proves the first part of the claim.

Similar to the proof of Proposition 3.4, the assumption on the uniformity of the differentiability and the norm definition in $D(\mathbb{R}^n)$ allows to prove the converse. \qed

The following result gives a criterion providing differentiability of the embedded map and generalizes the result for univariate set-valued maps in [36, Theorem 3.2.2].

**Proposition 3.7** (Characterization of Fréchet-differentiability). Let $F : \mathbb{R}^n \Rightarrow \mathbb{R}^n$ be given with images in $C(\mathbb{R}^n)$ and set $\overline{F}(x) = J_n(F(x))$.

(i) If $\overline{F}$ is Fréchet-differentiable at $x$, then $\delta^*(\eta,Y(l,F(\cdot)))$ is Fréchet-differentiable at $x$ uniformly in $\eta, l \in S^{n-1}$ with $\eta = l$ or $\eta \perp l$. For $n = 2$ this is also sufficient: if $\delta^*(\eta,Y(l,F(\cdot)))$ is Fréchet-differentiable at $x$ uniformly in $\eta, l \in S^{n-1}$ with $\eta = l$ or $\eta \perp l$, then $\overline{F}$ is Fréchet-differentiable at $x$.

(ii) $\overline{F}(\cdot)$ is (continuously) partially differentiable at $x$ if and only if $\delta^*(\eta,Y(l,F(\cdot)))$ is (continuously) partially differentiable at $x$ uniformly in $\eta, l \in S^{n-1}$ with $\eta = l$ or $\eta \perp l$.

**Proof.** (i) $\Rightarrow$: By Proposition 3.6, $f_n(\cdot; l) = \delta^*(l,F(\cdot)) = \delta^*(l,Y(l,F(\cdot)))$ is Fréchet-differentiable at $x$ uniformly in $l \in S^{n-1}$.

The same argument shows that $\overline{F}_{n-1}(\cdot; l)$ and hence,
\[
f_{n-1}(\cdot; \tilde{\eta}, l) = \delta^*(\tilde{\eta},\pi_{n-1}R^l_n Y(l,F(\cdot))) = \delta^*((R^l_n)^\top \pi_{n-1}^\top \tilde{\eta}, Y(l,F(\cdot)))
\]
is Fréchet-differentiable at $x$ uniformly in $l \in S^{n-1}$ and $\tilde{\eta} \in S^{n-2}$.

The mapping $\tilde{\eta} = (R^l_n)^\top \pi_{n-1}^\top \tilde{\eta}$ is a linear bijective isometry from $\mathbb{R}^{n-1}$ to $\langle l \rangle^\perp$ (the orthogonal complement of the linear span of the vector $l$), since
\[
\langle (R^l_n)^\top \pi_{n-1}^\top \tilde{\eta}, l \rangle = \langle \tilde{\eta},\pi_{n-1}R^l_n l \rangle = \langle \tilde{\eta},\pi_{n-1} e^n \rangle = \langle \tilde{\eta}, 0 \rangle = 0,
\]
and $\| (R^l_n)^\top \pi_{n-1}^\top \tilde{\eta} \| = \| \pi_{n-1}^\top \tilde{\eta} \| = \| \tilde{\eta} \|$.

Let $\eta \perp l$, then $\eta \in \langle l \rangle^\perp$, $R^l_n \eta \in \langle e^n \rangle^\perp$ and
\[
R^l_n \eta = \sum_{j=1}^{n-1} (R^l_n \eta, e^j) e^j + \langle R^l_n \eta, e^n \rangle e^n.
\]
Setting \( \tilde{\eta} = \pi_{n-1} R^I \eta \), \( \tilde{e}^j = \pi_{n-1} (e^j) \), we get \( \{ \tilde{e}^j : j = 1, \ldots, n-1 \} = \mathbb{R}^{n-1} \) and

\[
(R_n^I)^\top \pi_{n-1} \tilde{\eta} = (R_n^I)^\top \pi_{n-1} \pi_{n-1} R_n^I \eta = (R_n^I)^\top \pi_{n-1} \sum_{j=1}^{n-1} (R_n^I \eta, e^j) \tilde{e}^j = (R_n^I)^\top \sum_{j=1}^{n-1} (R_n^I \eta, e^j) e^j = \sum_{j=1}^{n-1} (R_n^I \eta, e^j) (R_n^I)^\top e^j = \sum_{j=1}^{n-1} \langle \eta, (R_n^I)^\top e^j \rangle (R_n^I)^\top e^j + \langle \eta, (R_n^I)^\top e^n \rangle (R_n^I)^\top e^n = \eta
\]

Therefore, \( \delta^*(\eta, Y(l, F(x))) \) is Fréchet-differentiable in \( x \) uniformly in \( l, \eta \in S^{n-1} \) with \( \eta \perp l \).

\[\text{“=}\text{”}: \text{ For } \eta = l, \text{ the Fréchet-differentiability of }\]

\[
\delta^*(l, Y(l, F(x))) = \delta^*(l, F(x)) = f_n(x; l)
\]

in \( x \) uniformly for \( l \in S^{n-1} \) follows.

For a given \( \tilde{\eta} \in S^{n-2} \), we set \( \eta = (R_n^I)^\top \pi_{n-1} \tilde{\eta} \in \langle l \rangle^\perp \). The function

\[
f_{n-1}(x; \tilde{\eta}, l) = \delta^*(\tilde{\eta}, \pi_{n-1} R_n^I y(l, F(x))) = \delta^*((R_n^I)^\top \pi_{n-1} \tilde{\eta}, Y(l, F(x))) = \delta^*(\eta, Y(l, F(x)))
\]

is Fréchet-differentiable at \( x \) uniformly for \( l \in S^{n-1} \) and \( \tilde{\eta} \in S^{n-2} \) which follows from the same property of \( \delta^*(\eta, Y(l, F(\cdot))) \) holding uniformly in \( l, \eta \in S^{n-1} \) with \( \eta \perp l \).

For \( n = 2 \) this is equivalent to the Fréchet-differentiability of \( \vec{F}_2(x) \), since there are no further recursive components in \( \vec{F}_2(x; l) \).

(ii) Let us first consider the \( n \) functions \( \vec{F}_i(t) = \vec{F}(x + te^i) \in \mathcal{D}(\mathbb{R}^n) \), \( i = 1, \ldots, n \). Then, \( \vec{F}_i(\cdot) \) is differentiable at 0 if and only if \( t \mapsto \delta^*(\eta, Y(l, F(x + te^i))) \) is differentiable at 0 uniformly in \( \eta, l \in S^{n-1} \) by [36, Theorem 3.2.2]. Hence, we already have a characterization of partial differentiability of \( \vec{F}(\cdot) \).

\( \vec{F}_i : \mathbb{R} \to \mathcal{D}(\mathbb{R}^n) \) is continuously differentiable at 0 if and only if for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that for all \( s \) with \( |s| \leq \delta \) there exists \( h(s) > 0 \) such that for all \( h \in (0, h(s)] \) we have

\[
\| \frac{1}{h} (\vec{F}_i(s + h) - \vec{F}_i(s)) - \frac{1}{h} (\vec{F}_i(h) - \vec{F}_i(0)) \| \leq \epsilon.
\]

The following estimations hold due to the estimate for Demyanov’s distance in [7, Proposition 4.11]:

\[
\sup_{l \in S^{n-1}} \max_{\eta \in S^{n-1}} \left| \frac{1}{h} \left( \delta^*(\eta, Y(l, F_i(s + h))) - \delta^*(\eta, Y(l, F_i(s))) \right) \right| + \left| \frac{1}{h} \left( \delta^*(\eta, Y(l, F_i(h))) - \delta^*(\eta, Y(l, F_i(0))) \right) \right| = \sup_{l \in S^{n-1}} \max_{\eta \in S^{n-1}} \left| \delta^*(\eta, Y(l, F_i(s + h))) - \delta^*(\eta, Y(l, F_i(s))) \right|
\]

...
The estimation of the right-hand side by the left-hand side (without the factor $\sqrt{2}$) also follows similarly.

With the characterization of continuous differentiability proved above it follows that $\overline{F}_i(\cdot)$ is continuously differentiable at 0 if and only if $h \to \delta^*(\eta, Y(l, F(x + he^i)))$ is continuously differentiable at 0 uniformly in $l, \eta \in S^{n-1}$. Hence, $\overline{F}_i(\cdot)$ is continuously partially differentiable at $x$ if and only if $x \to \delta^*(\eta, Y(l, F(x)))$ is continuously partially differentiable at $x$ uniformly in $l, \eta \in S^{n-1}$.

For a discussion of related smoothness criteria used in set-valued numerical analysis see [6].

4. Computation of convex invariant sets by Newton’s method.

4.1. Concept. The fundamental problem considered in this paper is to find convex invariant sets for a discrete-time dynamical system of the form

$$x_{k+1} = g(x_k) \quad (k \in \mathbb{N}_0)$$  \hspace{1cm} (15)

for some $g : \mathbb{R}^n \to \mathbb{R}^n$, that is convex sets $X_* \subset \mathbb{R}^n$ such that

$$g(X_*) = X_*,$$  \hspace{1cm} (16)

where $g(X_*)$ is the set that consists of all images $g(x), x \in X_*$.

Gaining knowledge about the invariant sets of a dynamical system is often an important step in understanding it. The most simple form of an invariant set is a fixed point, that is a point $x_* \in \mathbb{R}^n$ with $g(x_*) = x_*$. A fixed point $x_*$ is said to be stable if through the choice of $x_0$ sequences of the form (15) can be forced to remain in any (arbitrarily small) neighborhood of $x_*$, and it is called asymptotically stable if there is a neighborhood of $x_*$ such that sequences starting in this neighborhood eventually converge to $x_*$. A fixed point that is not stable is called unstable. These notions of stability can be extended to more complicated invariant sets. Notoriously difficult to compute are hyperbolic invariant structures which are unstable both in forward and in backward time.

The approach taken to solve problem (16) is to reformulate it first as a fixed point problem for a map $\overline{G}$ in a Banach space of directed sets and then, straightforward, as a zero-finding problem which will be solved by Algorithm 4.3 using Newton’s method. Here lies an important difference to conventional methods as e.g. implemented in GAIO: our approach employs derivatives and therefore needs a “set-oriented calculus” as a basis. As we are going to employ the Banach space calculus of directed sets for this purpose, we need to make sure that requirements
stemming from this approach are satisfied. They are summarized in the following two assumptions.

**Assumption 1.** There is a non-empty set of “compatible” convex sets

\[ C_0 = \{ X \in C(\mathbb{R}^n) : g(X) \in C(\mathbb{R}^n) \} \]  
(17)

and a map \( \vec{G} : J_n(C_0) \to D(\mathbb{R}^n) \) with

\[ \vec{G}(\vec{X}) = J_n(g(V_n(\vec{X}))) \]  
(18)

such that the map \( \vec{F} : J_n(C_0) \to D(\mathbb{R}^n) \) given by

\[ \vec{F}(\vec{X}) = \vec{X} - \vec{G}(\vec{X}) \]  
(19)

is Fréchet-differentiable.

In order to use the formalism from Section 2.2 for the Newton iteration, we need to have subsets within \( J_n(C_0) \) that are open with respect to the Banach space topology. As \( J_n(C_0) \subset D(\mathbb{R}^n) \), but not necessarily \( \langle J_n(C_0) \rangle = D(\mathbb{R}^n) \) — here \( \langle J_n(C_0) \rangle \) is the closure of the linear span of \( J_n(C_0) \) — \( J_n(C_0) \) may not contain open sets with respect to the topology of \( D(\mathbb{R}^n) \). To overcome this problem, we define \( \mathcal{X} = \langle J_n(C_0) \rangle \). As \( \mathcal{X} \) is a closed subspace of \( D(\mathbb{R}^n) \), it becomes a Banach space when equipped with the same norm as \( D(\mathbb{R}^n) \). From here on, \( \mathcal{X} \) will be the Banach space the Newton iteration is performed in.

**Assumption 2.** There are an open convex set \( U_0 \subset J_n(C_0) \subset \mathcal{X} \) and \( \vec{X}_0 \in U_0 \) such that \( \vec{F} \) satisfies the Kantorovich conditions given in Equations (7) and (8).

By Theorem 2.7, Assumption 2 asserts the well-definedness of the Newton iteration. Assumption 1 asserts that we work with those convex sets whose images under \( g \) are also convex, and that the dependence of the image sets is differentiable. In most cases we expect that this assumption is very hard to check a priori. In practice, during the Newton iteration one can check whether \( \vec{X}_k \in J_n(C_0) \) by asserting the convexity of images under \( g \) in each step. In some cases, it may also be possible to guarantee Assumption 1 by analytical means. For this, see e. g. the results on the convex range of some quadratic functions and the convexity of the nonlinear image of small balls in [37, 38], and also studies of subclasses of invariant sets (ellipsoids or polyhedral sets) as in [10] and references therein.

**Remark 4.1.** If the starting set \( \vec{X}_0 \) is suitably chosen, i. e. so that Assumption 2 is satisfied, by Theorem 2.7 the Newton iteration stays in the cone \( \vec{C}(\mathbb{R}^n) \) of embedded convex sets and the definition of the map \( \vec{G} \) in Equation (18) is sufficient.

However, in order to also treat cases where fulfillment of the assumptions cannot be checked, it may be worthwhile to consider alternative definitions of the map \( \vec{G} \). In these cases one can try to construct a Fréchet-differentiable lift \( \vec{G} : \mathcal{X} \to \mathcal{X} \) of \( g \) satisfying

\[ \vec{G}(J_n(X_0)) = J_n(g(X_0)) \quad (X_0 \in C_0) \]  
(20)

that preserves the images of convex elements in \( C_0 \).

In some cases — see the first three examples in this section — it is possible to define a natural extension \( \vec{G} \) to \( D(\mathbb{R}^n) \), as e. g. for linear maps \( g(x) = Ax \) in Lemma 3.2. Otherwise, there are two general approaches for defining the extended map \( \vec{G} \):
(i) In the special case that images of convex sets under the function $g(\cdot)$ remain convex (essentially either $g : \mathbb{R} \to \mathbb{R}$ or $g$ is affine), one can work with the entire spaces $\mathcal{C}_0 = \mathcal{C}(\mathbb{R}^n)$, $\mathcal{X} = \mathcal{D}(\mathbb{R}^n)$ and define $\overrightarrow{G} : \mathcal{D}(\mathbb{R}^n) \to \mathcal{D}(\mathbb{R}^n)$ by

$$\overrightarrow{G}(\overrightarrow{X}) = J_n(g(\co V_n(\overrightarrow{X}))) \quad (\overrightarrow{X} \in \mathcal{X}),$$

(21)

where $\co(A)$ denotes the convex hull of a set $A \subset \mathbb{R}^n$. The extension property (20) holds, since for $X_0 \in \mathcal{C}_0$

$$\overrightarrow{G}(J_n(X_0)) = J_n(g(\co V_n(J_n(X_0)))) = J_n(g(\co X_0)) = J_n(g(X_0))$$

as the visualization $V_n$ maps the embedded set $J_n(X_0)$ back to the convex set $X_0$ (see [8, Proposition 3.8(i)]) and the image $g(X_0)$ is convex by assumption.

(ii) If the condition in (i) does not hold for all convex sets, one can replace (17) by

$$\co g(X_0) \in \mathcal{C}_0 \quad (X_0 \in \mathcal{C}_0)$$

(22)

and define $\overrightarrow{G} : \mathcal{X} \to \mathcal{X}$ by

$$\overrightarrow{G}(\overrightarrow{X}) = J_n(\co g(V_n(\overrightarrow{X}))) \quad (\overrightarrow{X} \in \mathcal{X}).$$

(23)

As the invariant set $X_*$ is convex, the extension property (20) holds for $X_*:

$$\overrightarrow{G}(J_n(X_*)) = J_n(\co g(V_n(J_n(X_*)))) = J_n(\co g(X_*))$$

$$= J_n(\co X_*) = J_n(X_*)$$

(24)

In general, this approach can be defined even in the case that the convexity of the images cannot be guaranteed during the Newton iteration. If additionally the assumption (17) holds, the equality (24) above holds for all sets $X_0 \in \mathcal{C}_0$ so that the extension property (20) is preserved and both approaches (21) and (23) coincide yielding the same extension map $\overrightarrow{G}$.

Choice (i) has the advantage that one does not need to take care of the domain of $\overrightarrow{G}$ during the Newton iteration (since $\overrightarrow{G}$ is defined for all directed sets). However, a clear disadvantage of both approaches (i) and (ii) in comparison with a “natural” extension of $g$ is the use of the visualization mapping. When applied to directed sets with a mixed-type part, this might create nonconvex results, which makes it necessary to convexify the set or its image. This usually creates nonsmooth behavior of the map $\overrightarrow{G}$.

The equation for the invariant set can be stated as as a fixed point problem for directed sets

$$\overrightarrow{G}(\overrightarrow{X}_*) = \overrightarrow{X}_*$$

(25)

with the additional condition $\overrightarrow{X}_* = J_n(X_*)$, and subsequently as a zero-finding problem

$$\overrightarrow{F}(\overrightarrow{X}_*) = \overrightarrow{X}_* - \overrightarrow{G}(\overrightarrow{X}_*) = 0.$$  

(26)

**Lemma 4.2.** Let Assumption 1 and Assumption 2 be satisfied. If $\overrightarrow{X}_* \in \mathcal{U}_0$ satisfies (25) then $V_n(\overrightarrow{X}_*)$ is a convex invariant set satisfying (16).
Proof. The extension property (20) guarantees that a zero (set) \( X^* = J_n(X_*) \) satisfies additionally

\[
J_n(g(X_*)) = \overrightarrow{G}(J_n(X_*)) = J_n(X_*),
\]

by the result in [7, Theorem 4.17(iii)]. With the help of the visualization, the convex invariant set is regained by [8, Proposition 3.8(i)] via

\[
V_n(\overrightarrow{X_*}) = V_n(J_n(X_*)) = X_*.
\]

We now formulate Newton’s method for set-valued zero problems under Assumption 1 and Assumption 2.

**Algorithm 4.3** (Set-valued Newton’s method).

(i) For \( k = 0 \), choose a starting set \( X_0 \in C_0 \) with \( J_n(X_0) \in U_0 \) and set \( X_0 = J_n(X_0) \).

(ii) To compute \( X_{k+1} \) from \( X_k \), solve the operator equation

\[
\overrightarrow{F}'(\overrightarrow{X}_k)\Delta \overrightarrow{X}_k = -\overrightarrow{F}(\overrightarrow{X}_k)
\]

for \( \Delta X_k \) and set

\[
\overrightarrow{X}_{k+1} = \overrightarrow{X}_k + \Delta \overrightarrow{X}_k.
\]

(iii) Repeat step (ii) until the chosen stopping criterion (e.g. small norm of residual \( \| \overrightarrow{F}(\overrightarrow{X}_{k+1}) \| \leq \varepsilon \), small distance of subsequent iterates \( \| \overrightarrow{X}_{k+1} - \overrightarrow{X}_k \| \leq \varepsilon \) or maximum number of iterations reached) is satisfied.

Applying Theorem 2.7 with the setting \( U_0 = U_0, U = X = X \), we obtain the following convergence result for Algorithm 4.3.

**Proposition 4.4.** If a function \( g : \mathbb{R}^n \to \mathbb{R}^n \) and \( \overrightarrow{G}, \overrightarrow{F}, U_0 \) and \( \overrightarrow{X}_0 \) as in Assumptions 1 and 2 satisfy these assumptions, then the Newton iteration (27),(28) described in Algorithm 4.3 is well-defined and converges quadratically against the embedded invariant set \( X_* \), i.e.

\[
\|X_* - X_{k+1}\| \leq \frac{1}{2} L_k \|X_* - X_k\|^2
\]

(see Theorem 2.7 for details and notations).

Performing a Newton iteration in order to find zeroes of \( \overrightarrow{F} \), in each step we will have to solve an equation of the form

\[
\overrightarrow{F}'(\overrightarrow{X}_k) \cdot \Delta \overrightarrow{X}_k = -\overrightarrow{F}(\overrightarrow{X}_k).
\]

In the following sections, we will examine this problem in several contexts.
4.2. **Analytical example.**

**Example 4.5.** As the simplest example we consider a discrete-time linear dynamical system $g : \mathbb{R}^n \to \mathbb{R}^n$ given by $g(x) = Ax$, where $A \in \mathbb{R}^{n \times n}$ is an invertible matrix with the spectrum satisfying $\sigma(A) \cap \{z \in \mathbb{C} : |z| = 1\} = \emptyset$. Its only invariant compact set is $\{0\}$.

The map $g$ clearly maps convex sets to convex sets and can thus be lifted to a linear map $\overrightarrow{G} : \mathcal{D}(\mathbb{R}^n) \to \mathcal{D}(\mathbb{R}^n)$ with $\overrightarrow{G}(\overrightarrow{X}) := A\overrightarrow{X}$ in the way described in Lemma 3.2. With the notations of Algorithm 4.3 we set $C_0 = \mathcal{C}(\mathbb{R}^n)$, $\mathcal{X} = \mathcal{D}(\mathbb{R}^n)$.

A simple calculation shows that the extension property (20) holds for this natural extension $\overrightarrow{G}$, since $\overrightarrow{GJ}_n(\overrightarrow{X}_0) = J_n(A\overrightarrow{X}_0)$. We see that $\overrightarrow{F}$ is also a linear map and therefore equals to its derivative, i.e.

$$\overrightarrow{F}'(\overrightarrow{X})\overrightarrow{W} = \overrightarrow{W} - A\overrightarrow{W}.$$  

Starting from an arbitrary directed set $\overrightarrow{X}_0 \in \mathcal{D}(\mathbb{R}^n)$, we therefore obtain for the first step of the Newton iteration

$$\overrightarrow{X}_1 = \overrightarrow{X}_0 + \Delta\overrightarrow{X}_0,$$

where $\Delta\overrightarrow{X}_0$ solves

$$\overrightarrow{F}'(\overrightarrow{X}_0)\Delta\overrightarrow{X}_0 = -\overrightarrow{F}(\overrightarrow{X}_0) \quad (30)$$

$$\iff \Delta\overrightarrow{X}_0 - A\Delta\overrightarrow{X}_0 = -(\overrightarrow{X}_0 - A\overrightarrow{X}_0) \quad (31)$$

Clearly, $\Delta\overrightarrow{X}_0 = -\overrightarrow{X}_0$ is a solution. To see that it is the only solution, observe that $\overrightarrow{F}'(\overrightarrow{X}_0)$ is an injective operator, as

$$\overrightarrow{F}'(\overrightarrow{X}_0)\overrightarrow{X} = 0 \iff A\overrightarrow{X} = \overrightarrow{X}$$

which is satisfied for $X = \{0\}$ only, i.e. for $\overrightarrow{X} = 0$ only. Therefore $\overrightarrow{X}_1 = 0 = J_n(\{0\})$, and the Newton iteration converges after one step.

Already in this simple example it is difficult to check the Kantorovich conditions given in Equations (7) and (8). In particular, it is not obvious that $\overrightarrow{F}'$ is invertible. However, for the even simpler case $A = \alpha I$, $\alpha \notin \{0, 1\}$, the convergence assumptions from Proposition 4.4 can be shown. Since $\overrightarrow{F}'(\overrightarrow{X})^{-1}\overrightarrow{Z} = \frac{1}{1-\alpha}\overrightarrow{Z}$ and the derivative (and its inverse) does not depend on the set $\overrightarrow{X}$ itself, condition (7) is satisfied even with $K = 0$. The constant $\eta$ equals the norm of $\overrightarrow{X}_0$, so that condition (8) is fulfilled, which yields quadratic convergence.

This result is of course hardly surprising. However, the example demonstrates an important feature of our approach. We have seen that in this example, the Newton iteration behaves completely independently of the eigenvalues of $A$. In contrast, a classical fixed point iteration would depend very sensitively on the eigenvalues of $A$. It would govern the rate of convergence (see e.g. [12], [34]), and, even more important, the iteration would converge only for $\rho(A) < 1$, that is, in the case where the invariant set is attractive. It is a crucial advantage of our approach that the Newton iteration works regardless of the stability properties of the set in question.

We note that this property of Newton’s method even extends to truly hyperbolic invariant structures, which one has no chance to compute by fixed set iterations.
5. Numerical realization of the set-valued Newton’s method.

5.1. Approximation of directed sets. A question concerning directed sets that arises naturally in application-oriented contexts is that of the finite-dimensional approximation of directed sets. Given that an element of $D(\mathbb{R}^n)$ is essentially a map $S^{n-1} \to \mathbb{R} \times D(\mathbb{R}^{n-1})$, a natural idea to approximate it is to replace $S^{k-1}$ ($k = 1, \ldots, n - 1$) by a finite number of unit vectors and recursively prescribe the values of both components on these directions. In the following we document that in this way one actually obtains an approximation result.

**Definition 5.1.** Let $\widetilde{A}, \widetilde{B} \in D(\mathbb{R}^n)$. The visual distance between $\widetilde{A}$ and $\widetilde{B}$ is

$$\delta_V(\widetilde{A}, \widetilde{B}) := d_H(V_n(\widetilde{A}), V_n(\widetilde{B})),$$

where $d_H$ is the Hausdorff distance of compact sets.

From the metric properties of the Hausdorff distance, one can easily prove the following result.

**Lemma 5.2.** (i) $\delta_V$ is a pseudometric on $D(\mathbb{R}^n)$.

(ii) The pseudometric is zero if and only if the visualizations of both directed sets coincide. E.g. convex compact sets which are symmetric to the origin have a distance zero to their inverse.

(iii) In particular, restricted to the cone $\widetilde{C}(\mathbb{R}^n)$ of embedded convex, compact sets, $\delta_V$ is a metric.

**Remark 5.3.** The visual distance is weaker than the distance $d_V(C, D) = \|J_n(C) - J_n(D)\|$ on the cone $C(\mathbb{R}^n)$ of convex compact sets that is introduced in [7, Definition 4.8], since the latter metric is equivalent to the stronger Demyanov metric. Therefore the visual distance is a lower bound on the metric induced by the norm in $D(\mathbb{R}^n)$ for embedded sets.

We can now state two approximation results for convex sets which are formulated with respect to the visual distance.

**Lemma 5.4.** Let $G^n = \{l^i \in S^{n-1} | i = 1, \ldots, m\}$. For $C \in C(\mathbb{R}^n)$ and $l^i \in G^n$ let $y(l^i, C) \in Y(l^i, C)$ be arbitrary supporting points, and define

$$\widetilde{C}^m := \text{co} \{y(l^i, C) | i = 1, \ldots, m\} \text{ and}$$

$$C^m := \bigcap_{i=1,\ldots,m} \{x \in \mathbb{R}^n | \langle l^i, x \rangle \leq \delta^*(l^i, C)\}.$$

Let $c \in C$ be arbitrary. If $d_H(G^n, S^{n-1}) \leq \varepsilon$ for some $\varepsilon > 0$, then

$$\delta_V(J_n(C), J_n(\widetilde{C}^m)) \leq 2 \cdot \|C - c\| \cdot \varepsilon,$$

$$\delta_V(J_n(C), J_n(C^m)) \leq 3 \cdot \|C - c\| \cdot \varepsilon,$$

where $\|C\| = \max_{c \in C} \|c\|$. The second inequality is only valid for $\varepsilon \leq \frac{1}{2}$ and $C^m$ being bounded.

**Proof.** Apply Proposition 3.4 in [2] to the set $C - c$. 

This lemma implies that good approximations $\widetilde{C}^m$ and $C^m$ of a convex, compact set $C$ with respect to the visual distance are obtained for “evenly spread” discretizations $G^n$ of the unit sphere. For $n \geq 2$, the number $m$ of elements in $G^n$ necessary to obtain $d_H(G^n, S^{n-1}) \leq \varepsilon$ is roughly proportional to $\varepsilon^{-1-n}$. This allows a first
estimation of the computational effort necessary to perform this method in higher dimensional spaces.

5.2. Realization of the Newton step. We now consider how the set-valued Newton iteration can be performed numerically. We first consider the one-dimensional case, for which no discretization is needed, as \( D(\mathbb{R}) \) is two-dimensional.

\( \text{a) Realization for directed intervals.} \) In this case, from the general scheme described in Subsec. 4.1 we obtain a map \( \widetilde{F} : U_0 \to X, U_0 \subset X \subset D(\mathbb{R}) \), for which we want to solve the equation \( \widetilde{F}(\overline{x}) = 0 \).

Writing \( \overline{x} = (x_-, x_+)^\top \) and \( \Delta \overline{x} = (\Delta x_-, \Delta x_+)^\top \), the map \( \widetilde{F} \) can be written as
\[
\widetilde{F}(\overline{x}) = [f_1(x_-, x_+), f_2(x_-, x_+)]
\]
with two functions \( f_i(\cdot, \cdot), i = 1, 2 \), so that the corresponding linear equation reads
\[
\begin{bmatrix}
\partial_1 f_1(x_-, x_+) & \partial_2 f_1(x_-, x_+)
\end{bmatrix}
\begin{bmatrix}
\Delta x_-
\end{bmatrix}
- \begin{bmatrix}
\Delta x_+
\end{bmatrix}
= - \begin{bmatrix}
\partial_1 f_1(x_-, x_+), & \partial_2 f_1(x_-, x_+)
\end{bmatrix}
\begin{bmatrix}
(f_1(x_-, x_+), f_2(x_-, x_+))
\end{bmatrix},
\]
where \( \partial_1 \) and \( \partial_2 \) denote the partial differentiation w.r.t. the first and second argument, respectively.

Example 5.5. Numerically integrating the one-dimensional ODE
\[
x'(t) = x(t)(1 - x(t))
\]
with the explicit Euler method with step size \( h \) leads to the map \( g : \mathbb{R} \to \mathbb{R}, g(x) = x + h \cdot x \cdot (1 - x) \). Setting \( h = \frac{1}{12}, \) we want to find intervals \([x_-, x_+]\) with end points \( x_-, x_+ \in \mathbb{R} \) with \( x_+ \geq x_- \) such that
\[
g([x_-, x_+]) = [x_-, x_+].
\]
We have \( g'(x) = \frac{11}{12} - x \), so \( g \) is strictly increasing on the interval \((-\infty, \frac{11}{2}], \) and equation (33) has three different solutions that are contained in this interval:
\[
\{0\}, \quad \text{i.e. } [x_-, x_+] = [0, 0],
\{1\}, \quad \text{i.e. } [x_-, x_+] = [1, 1],
\{0, 1\}, \quad \text{i.e. } [x_-, x_+] = [0, 1].
\]
For Algorithm 4.3 we set \( C_0 = \{[x_-, x_+] : x_-, x_+ \in [-5, 5], x_- \leq x_+\}, X = D(\mathbb{R}). \)

With the natural extension \( \overline{G} \) defined as
\[
\overline{G}(\overline{x}) = [g(x), g(x_+)] \quad \text{for } \overline{x} = [x_-, x_+],
\]
the extension property (20) can be shown due to the monotonicity of \( g(\cdot) \) in \([-5, 5] \subset (-\infty, \frac{11}{2}] \):
\[
\overline{G}(J_1([x_-, x_+])) = [g(x), g(x_+)] = J_1([g(x), g(x_+))] = J_1(g([x_-, x_+])),
\]
where \( x_-, x_+ \in \mathbb{R} \) with \( x_- \leq x_+ \). We obtain
\[
\overline{F}(\overline{x}) = \overline{x} - g(\overline{x}) = [x_- - g(x), x_+ - g(x_+)]
\]
and from Corollary 3.5 we see that
\[
\overline{F}^i(\overline{x}) w = w_- \cdot \partial_1 \overline{F}(\overline{x}) + w_+ \cdot \partial_2 \overline{F}(\overline{x}),
\]
where
\[
\partial_1 \overline{F}(\overline{x}) = [1 - g'(x_+), 0], \quad \partial_2 \overline{F}(\overline{x}) = [0, 1 - g'(x_+)].
\]
and $\overrightarrow{w} = [w^-, w^+]$. The Newton step of the zero-finding problem $\overrightarrow{F}(\overrightarrow{X}) = 0$ is given by

$$\overrightarrow{F}'(\overrightarrow{X}_k)\Delta \overrightarrow{X}_k = -\overrightarrow{F}(\overrightarrow{X}_k),$$

where $\overrightarrow{X}_k = [x_k^-, x_k^+]$, $\Delta \overrightarrow{X}_k = [\Delta x_k^-, \Delta x_k^+] \in D(\mathbb{R})$. The equivalent linear system of dimension $2 \times 2$ is

$$\begin{pmatrix}
1 - g'(x_k^-) & 0 \\
0 & 1 - g'(x_k^+) \\
\end{pmatrix}
\begin{pmatrix}
\Delta x_k^- \\
\Delta x_k^+ \\
\end{pmatrix}
= \begin{pmatrix}
-x_k^- + g(x_k^-) \\
-x_k^+ + g(x_k^+) \\
\end{pmatrix}.$$

(34)

It should be noted that here the derivative $\overrightarrow{F}'(\overrightarrow{X}_k)$ becomes singular, if $g'(x_k^-) = 1$ or $g'(x_k^+) = 1$, i.e. for $x_k^- = \frac{1}{2}$ or $x_k^+ = \frac{1}{2}$. In this case, equation (34) has no solution.

The Kantorovich condition (7) is satisfied, since $g'()$ is Lipschitz as long as the inverse of the matrix in (34) exists for the starting set $\overrightarrow{X}_0$. The condition (8) is not satisfied for the chosen starting intervals in Tables 1 and 2, but only for suitable starting intervals closer to the corresponding solutions, e.g. for $\overrightarrow{X}_0 \in \{[0.9, 1.1], [-0.1, 1.1], [-0.1, 0.1]\}$ so that quadratic convergence can be expected near all invariant intervals.

| $k$ | $\overrightarrow{X}_k$ | $\Delta \overrightarrow{X}_k$ | $\overrightarrow{F}(\overrightarrow{X}_k)$ |
|-----|------------------------|-------------------------------|---------------------------------|
| 0   | $[0.7000000000, 1.100000000]$ | $0.525000, -0.091667$ | $[-0.021000, 0.011000]$ |
| 1   | $[1.225000000, 1.008333333]$ | $-0.190086, -0.008265$ | $[0.027562, 0.000840]$ |
| 2   | $[1.034913793, 1.000068306]$ | $-0.033774, -0.000608$ | $[0.003613, 0.000007]$ |
| 3   | $[1.001139411, 1.000000000]$ | $-0.001138, -0.000000$ | $[0.000114, 0.000000]$ |
| 4   | $[1.000001295, 1.000000000]$ | $-0.000000, 0.000000$ | $[0.000000, -0.000000]$ |
| 5   | $[1.000000000, 1.000000000]$ | $-0.000000, 0.000000$ | $[0.000000, -0.000000]$ |

Table 1. Example 5.5: convergence of Newton’s iteration for $X_0 = [0.7, 1.1]$

In Tables 1 and 2 we document two illustrative iterations for this system. Table 1 describes a case, where Newton’s method creates improper intervals $\overrightarrow{X}_k$ and $\overrightarrow{F}(\overrightarrow{X}_k)$ for $k \geq 1$, e.g. $\overrightarrow{X}_1 \approx [1.225000, 1.008333]$, $\overrightarrow{F}(\overrightarrow{X}_1) = [0.027562, 0.000840]$. Nevertheless, the convergence of the directed intervals $\overrightarrow{X}_k$ to the point $X = \{1\}$ can be easily recognized. In Table 2, the Newton iteration converges to $X = [0.0, 1.0]$ (from the outside at $k \geq 1$). In both cases, quadratic convergence can be observed.

There is a history of Newton iterations using interval arithmetic in order to compute (typically tiny) intervals guaranteed to contain precisely one zero of a function, see e.g. [30, 31, 19]. It is worthwhile to note that Example 5.5 also illustrates a difference between these methods and the Newton method using directed sets. While the result in this example is also a singleton interval, an iteration using interval arithmetic could not have produced the sequence of iterates generated in the example, since it contains “improper intervals”, i.e. additive inverses of representations of intervals as directed sets.
In actual implementations we replace \( g \) with a set \( \mathcal{C} \) of \( n \) disjoint sets \( \mathcal{C}_l \) which approximate \( g(X) \) around \( X = \{ \alpha \in \mathcal{C}_l : \langle \gamma, x \rangle = 0 \} \) for \( \gamma \) in a directed set \( \mathcal{D} \) of \( \mathbb{R}^n \). Since for \( X \in \mathcal{C}_0 \) and the approximation \( g(X) \) only the variables \( x = (\delta^*(l^1, X), \ldots, \delta^*(l^m, X)) \) matter, we consider the functions \( F_m, G_m : \mathbb{R}^m \rightarrow \mathbb{R}^m \) with

\[
G_m(x) = (\delta^*(l^1, g_m(X)), \ldots, \delta^*(l^m, g_m(X))),
\]

\[
F_m(x) = x - G_m(x)
\]

as a discretization for the extension map \( \overline{G} \) and of \( \overline{F} \).

As the convex set \( X \) is approximated only by \( x = (\delta^*(l^1, X), \ldots, \delta^*(l^m, X)) \), we allow arbitrary vectors \( x = (x_1, \ldots, x_m) \) in the Newton iteration, thus ignoring the lower-dimensional parts \( \overline{X}_{n-1}(l^i), i = 1, \ldots, m \), in a directed set \( \overline{X} \).

The discretized equation for the Newton step

\[
F_m(x_k) \Delta x_k = -F_m(x_k) \quad (x_k, \Delta x_k \in \mathbb{R}^m)
\]
and the corresponding directed sets based on \( m \) directions are visualized. If no knowledge about the derivative in (37) is available, we approximate the derivative \( F_m' \) using finite differences.

6. Numerical examples. In the following we present numerical examples for the efficacy of the method in a two-dimensional setting. In each case, the invariant sets to be computed are stable either in forward or in backward time. However, we note that these examples can easily be extended to higher-dimensional systems possessing invariant sets that are unstable both in forward and in backward time. One only has to add a third dimension e. g. with linear dynamics independent from the first two variables, and stability just opposite of that of the invariant set in two dimensions. So e. g. in the case of Example 6.2(i) with \( \alpha = 0.1 \), extending the map \( g \) by a third component \( g_3(x_1, x_2, x_3) = 2x_3 \) makes the invariant set (which is contained in the plane \( x_3 = 0 \)) unstable in both forward and backward time.

Example 6.1. For \( \alpha > 0 \) and irrational \( \delta \in (0, 2\pi) \) we define \( \tilde{g} : \mathbb{R} \rightarrow \mathbb{R} \) and \( g : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) by

\[
\tilde{g}(r) = \alpha \cdot \left( 1 + \frac{\arctan(r - \alpha)}{\arctan(\alpha)} \right) \quad \text{and} \quad g(v) = \tilde{g}(\|v\|_2) \cdot \left( \begin{array}{c} \cos(\delta) \\ -\sin(\delta) \end{array} \right) .
\]

Thus \( g \) maps a circle with radius \( r \) centered at the origin to a circle with radius \( \tilde{g}(r) \) also centered at the origin, while on the circle, all points are rotated by an angle \( \delta \). As \( \tilde{g} \) has unstable fixed points at \( r = 0 \) and \( r = 2\alpha \) and a stable fixed point at \( r = \alpha \), and the rotation by \( \delta \) has no periodic points on each circle, there are precisely three invariant sets for \( g \), namely the balls with these radii centered at the origin.

In this example, we observe convergence of the two-dimensional discretization following Equation (37). Depending on the choice for an initial guess, one can create sequences converging to any of the three invariant sets.

More precisely, we set \( \alpha = 1 \), discretize \( S^1 \) by \( m = 100 \) evenly spaced unit vectors and start Newton iterations with discretizations of four different ellipses. Starting (a) with an ellipse centered in the point \((2, 2, 2)\), with semimajor axes of lengths 6.2 and 6.5, respectively, aligned to the coordinate axes, the iteration converges to the invariant set \( B_3(0) \). The set \( B_1(0) \) is approached by two iterations starting with ellipses (b) centered in the point \((0.1, 0)\), with semimajor axes of lengths 1.1 and 1.3, respectively, tilted by an angle of 0.3; and (c) centered in the origin, with semimajor axes of lengths 0.8 and 1.3, respectively, tilted by an angle of 0.3. Starting (d) with an ellipse centered in the origin, with semimajor axes 0.05 and 0.01, respectively, the iteration converges to \( \{0\} \). Figure 2 illustrates the first three iterations, the numerical results of all four are given in Table 3. Figure 3 displays the differences \( X_1 - X_0 \) for the iterations (a) and (c). It is worthwhile to note that convergence to the invariant set consisting of the origin is slower that in the other two cases, and one observes that condition numbers of \( F_m(x_k) \) increase during the iteration. The reason for this might be that the origin is on the boundary of the convex cone.

Example 6.2. (i) Let \( g : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \),

\[
g\left( \begin{array}{c} x_1 \\ x_2 \end{array} \right) = \left( \begin{array}{cc} (1 + \alpha) & 0.15625 \\ 0.064 & (1 + \alpha) \end{array} \right) \left( \begin{array}{c} x_1 \\ x_2 \end{array} \right) - \alpha \cdot \left( 0.64x_1^2 + 1.5625x_2^2 \right) \left( \begin{array}{c} x_1 \\ x_2 \end{array} \right)
\]

with \( \alpha \in \mathbb{R} \). We consider the values \( \alpha = 0.1 \) and \( \alpha = -0.1 \). In the first case there exists a convex, compact stable invariant set while in the second case
Table 3. Norm of the values of $F$ as defined by Equation (36) for the Newton iterations for Example 6.1. Convergence to the two invariant sets $B_2(0)$ (Case (a)) and $B_1(0)$ (Cases (b) and (c)) is quadratic, while the singleton set $\{0\}$ is approached only linearly (Case (d)).

| Case (a) | Case (b) | Case (c) | Case (d) |
|----------|----------|----------|----------|
| $k$ | $\|F(x_k)\|$ |
| 0 | 42.22294371 | 3.87222350 | 0.00009537 | 0.00000027 |
| 1 | 3.2227074 | 0.18432471 | 0.00000034 | 0.00000024 |
| 2 | 0.48549194 | 0.00963379 | 0.00000003 | 0.00000024 |
| 3 | 0.03440850 | 0.00244721 | 0.00000003 | 0.00000024 |
| 4 | 0.00035687 | 0.00009537 | $< 10^{-12}$ | $< 10^{-12}$ |
| 5 | 0.00000027 | 0.00000024 | $< 10^{-12}$ | 0.00031239 |

there exists a convex, compact unstable invariant set (hyperbolic dynamical system with the origin as equilibrium which is a nodal source, since the eigenvalues of $g'(0,0)$ are $\frac{6}{5}$ and 1 for $\alpha = 0.1$ resp. 1 and $\frac{1}{5}$ for $\alpha = -0.1$). Both invariant sets form ellipses centered at the origin. Table 4 and Figure 4(a)–(b) show the Newton iterations for the corresponding function $F_m$ as defined by (36). Note, in particular, that Newton’s method allows to easily approximate the unstable set for $\alpha = -0.1$.

(ii) We modify the map $g : \mathbb{R}^2 \to \mathbb{R}^2$ from (i) as follows:

$$
g(x_1, x_2) = \begin{pmatrix} 0.8 & 0.796875 \\ 0.1792 & 1.4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 0.1 \cdot \begin{pmatrix} (0.64x_1^2 + 2.1875x_2^2)x_1 \\ (0.64x_1^2 + 1.40625x_2^2)x_2 \end{pmatrix}
$$

Again this map has a convex, compact invariant set but now it has no simple geometric form as in the previous case (hyperbolic dynamical system with the origin as equilibrium which is a nodal source, since the eigenvalues of $A$ are approximately 1.58 and 0.61). Nevertheless, we are still able to compute this set using Newton’s method as shown in Table 4 and Figure 4(c)–(d). For comparison we also computed a box covering of the invariant set of $g$ using the subdivision algorithm presented in [12] for the computation of unstable manifolds and global attractors. As shown in Figure 4(d), the results agree very well. On the other hand the number of evaluations of the function $g$ needed for the subdivision algorithm was more than 6 times higher than for the Newton iteration.

7. Conclusion. In this article, a novel approach to the computation of invariant sets of dynamical systems was presented. Unlike conventional methods, the set-oriented approach followed here employs derivatives and is formulated analogously to the point-wise case. Using the Banach space formalism of directed sets for this purpose, we were able to reformulate the invariance condition as a zero-finding problem in a suitable Banach space. In the reformulation either natural extension maps $\mathcal{G}$ were applied (Examples 4.5, 5.5 and 6.1) or the approach as in Sec. 5.2 was followed (Example 6.2).

It was shown that Newton iterations for the solution of this problem can be employed to approximate compact convex invariant sets of dynamical systems regardless of their dynamical stability. Under this approach, stable and unstable
Table 4. Norm of the function of $F$ as defined by Equation (36) for the Newton iterates of Example 6.2.

| $\alpha$ | $F(x_k)$ | $F(x_k)$ | $F(x_k)$ |
|----------|----------|----------|----------|
| $k$      |          |          |          |
| 0        | 1.636178218677 | 0.623666707123 | 3.019902689441 |
| 1        | 0.413776747887  | 0.191253669155  | 0.537252706487 |
| 2        | 0.083250696241  | 0.031418837852  | 0.176229318237 |
| 3        | 0.010136637687  | 0.001443154128  | 0.088586976802 |
| 4        | 0.000885036260  | 0.000026775615  | 0.036522553864 |
| 5        | 0.00015968827   | 0.000000029676  | 0.015513988506 |
| 6        | 0.000000010704  | 0.000000000000  | 0.001157750356 |
| 7        | 0.000000000000  | 0.000014696139  |          |
| 8        | 0.000000011695  |                    |          |
| 9        | 0.000000000000  |                    |          |

Invariant sets behave in principle in an identical way, and quadratic order of convergence can be proved or observed independently of the dynamical behavior of the system.

The errors introduced in the numerical approach in this part of the subsection to discretize directed sets by $m$ directions and to approximate the image $g(X_m)$ by the evaluation of test points need to be studied in more detail. Especially the discretization of the Newton step (29) with finitely many directions of the two equations

$$\overrightarrow{D}_{n-1}(\overrightarrow{x}; l) = -\overrightarrow{F}_{n-1}(\overrightarrow{x}; l) \quad (l \in S^{n-1}),$$

$$d_{n}(\overrightarrow{x}; l) = -f_{n}(\overrightarrow{x}; l) \quad (l \in S^{n-1}),$$

where $\overrightarrow{D}(\overrightarrow{x}) = \overrightarrow{F}'(\overrightarrow{x}) \Delta \overrightarrow{x}$ with the representation $(\overrightarrow{D}_{n-1}(\overrightarrow{x}; l), d_{n}(\overrightarrow{x}; l))_{l \in S^{n-1}},$ and the errors in the corresponding solutions should be investigated.

If the equations (38)–(39) are not solvable (e.g. if only a specific subclass of directed sets is considered), we can apply a Gauß-Newton method for the discretized finite-dimensional setting, see [5].

It is the authors’ belief that certain “trivial” types of nonconvex invariant sets, e.g. unions of several convex sets that are cyclically permuted, can also be computed with this approach by regarding suitable iterates of the map $g$. Furthermore, it is the author’s hope that on the basis of this paper, future work will relax some of the restrictions assumed in this work, and will extend the framework to allow the computation of more general invariant sets. We note that following examples given e.g. by Wang [46], the approach in this paper could be used to prove inverse function theorems for set-valued mappings.

REFERENCES

[1] J.-P. Aubin, Mutational and Morphological Analysis. Tools for Shape Evolution and Morphogenesis, Systems & Control: Foundations & Applications, Birkhäuser Boston Inc., Boston, MA, 1999.

[2] R. Baier, C. Bäskens, I. A. Chahma and M. Gerdts, Approximation of reachable sets by direct solution methods of optimal control problems, Optim. Meth. Softw., 22 (2007), 433–452.
THE COMPUTATION OF CONVEX INVARIANT SETS VIA NEWTON’S METHOD

[2] R. Baier and E. M. Farkhi, Directed derivatives of convex compact-valued mappings, in Advances in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory (1873–1950) (eds. N. Hadjisavvas and P. M. Pardalos), vol. 54 of Nonconvex Optimization and Its Applications, Kluwer Academic Publishers, Dordrecht-Boston-London, (2001), 501–514.

[3] R. Baier and E. M. Farkhi, The directed subdifferential of DC functions, in Nonlinear Analysis and Optimization II: Optimization. A Conference in Celebration of Alex Ioffe’s 70th and Simeon Reich’s 60th Birthdays, June 18–24, (2008), Haifa, Israel (eds. A. Leizarowitz, B. S. Mordukhovich, I. Shafrir and A. J. Zaslavski), vol. 513 of AMS Contemp. Math., AMS and Bar-Ilan University, (2010), 27–43, http://www.ams.org/bookstore-getitem/item=CONM-514.

[4] R. Baier and M. Hessel-von Molo, Newton’s method and secant method for set-valued mappings, in Proceedings on the 8th International Conference on “Large-Scale Scientific Computations” (LSSC 2011), June 6–10, 2011, Sozopol, Bulgaria (eds. I. Lirkov, S. Margenov and J. Wansiewski), vol. 7116 of Lecture Notes in Comput. Sci., Springer-Verlag, Berlin-Heidelberg, 2012, 91–98.

[5] R. Baier and G. Perria, Set-valued Hermite interpolation, J. Approx. Theory, 163 (2011), 1349–1372.

[6] R. Baier and E. M. Farkhi, Differences of convex compact sets in the space of directed sets. I. The space of directed sets, Set-Valued Anal., 9 (2001), 217–245.

[7] R. Baier and E. M. Farkhi, Differences of convex compact sets in the space of directed sets. II. Visualization of directed sets, Set-Valued Anal., 9 (2001), 247–272.

[8] W.-J. Beyn, The numerical computation of connecting orbits in dynamical systems, IMA J. Numer. Anal., 10 (1990), 379–405, http://imajna.oxfordjournals.org/cgi/content/abstract/10/3/379.

[9] F. Blanchini, Set invariance in control, Automatica, 35 (1999), 1747–1767.

[10] M. Dellnitz, G. Froyland and O. Junge, The algorithms behind GAIO—set oriented numerical methods for dynamical systems, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems (ed. B. Fiedler), Springer, (2001), 145–174.

[11] M. Dellnitz and A. Hohmann, A subdivision algorithm for the computation of unstable manifolds and global attractors, Numer. Math., 75 (1997), 293–317.

[12] M. Dellnitz and O. Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal., 36 (1999), 491–515.

[13] V. F. Demyanov and A. M. Rubinov, Constructive Nonsmooth Analysis, vol. 7 of Approximation and Optimization, Peter Lang, Frankfurt am Main–Berlin–Bern–New York–Paris–Wien, 1995, Updated and revised English edition of V. F. Demyanov, A. M. Rubinov, Foundations of Nonsmooth Analysis, and Quasidifferentiable Calculus, Optimizatsiya i Operatsii, 23, Nauka, Moscow, 1990 (in Russian).

[14] L. Dieci, J. Lorenz and R. D. Russell, Numerical calculation of invariant tori, SIAM J. Sci. Statist. Comput., 12 (1991), 607–647, http://link.aip.org/link/?SCE/12/607/1.

[15] N. S. Dimitrova and S. M. Markov, Interval methods of Newton type for nonlinear equations, Pliska Stud. Math. Bulg., 5 (1983), 105–117.

[16] T. Donchev and E. M. Farkhi, Fixed set iterations for relaxed Lipschitz multimaps, Nonlinear Anal., 53 (2003), 997–1015.

[17] H. Hadwiger, Minkowskische Addition und Subtraktion beliebiger Punktmengen und die Theoreme von Erhard Schmidt, Math. Z., 53 (1950), 210–218.

[18] E. Hansen, A multidimensional interval Newton method, Reliab. Comput., 12 (2006), 253–272.

[19] A. J. Homburg, H. M. Osinga and G. Vegter, On the computation of invariant manifolds of fixed points, ZAMM Z. Angew. Math. Phys., 46 (1995), 171–187.

[20] C. S. Hsu, Global analysis by cell mapping, Internat. J. Bifurc. Chaos Appl. Sci. Engrg., 2 (1992), 727–771.

[21] L. V. Kantorovich, On Newton’s method for functional equations, Dokl. Akad. Nauk SSSR, 59 (1948), 1237–1240.

[22] L. V. Kantorovich, The majorant principle and Newton’s method, Dokl. Akad. Nauk SSSR, 76 (1951), 17–20.

[23] L. V. Kantorovich and G. P. Akilov, Functional Analysis, 2nd edition, Pergamon Press, Oxford, 1982, Translated from the Russian by H. L. Silcock.
[25] I. G. Kevrekidis, R. Aris, L. D. Schmidt and S. Pelikan, Numerical computation of invariant circles of maps, *Phys. D*, **16** (1985), 243–251, [http://www.sciencedirect.com/science/article/B6TVK-46MKSBY-2/2/48df19752ce87d5170b0eabe206cbf1](http://www.sciencedirect.com/science/article/B6TVK-46MKSBY-2/2/48df19752ce87d5170b0eabe206cbf1).

[26] I. G. Kevrekidis, C. W. Gear and G. Hummer, Equation-free: The computer-aided analysis of complex multiscale systems, *AIChE Journ.*, **50** (2004), 1346–1354.

[27] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg and C. Theodoropoulos, Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis, *Commun. Math. Sci.*, **1** (2003), 715–762, [http://projecteuclid.org/getRecord?id=euclid.cms/1119655353](http://projecteuclid.org/getRecord?id=euclid.cms/1119655353).

[28] D. Klatte and B. Kummer, *Nonsmooth equations in optimization. Regularity, calculus, methods and applications*, vol. 60 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, 2002.

[29] B. Krauskopf, H. M. Osinga, E. J. Doedel, M. E. Henderson, J. Guckenheimer, A. Vladimirsky, M. Dellnitz and O. Junge, A survey of methods for computing (un)stable manifolds of vector fields, *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, **15** (2005), 763–791.

[30] R. Krawczyk, Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerschranken, *Computing (Arch. Elektron. Rechnen)*, **4** (1969), 187–201.

[31] R. Krawczyk, Einschließung von Nullstellen mit Hilfe einer Intervallarithmetik, *Computing (Arch. Elektron. Rechnen)*, **5** (1970), 356–370.

[32] E. Kreuzer, Analysis of chaotic systems using the cell mapping approach, *Arch. Appl. Mech.*, **55** (1985), 285–294.

[33] S. M. Markov, Some applications of extended interval arithmetic to interval iterations, *Comput. Suppl.*, **2** (1980), 69–84.

[34] K. Nickel, Das Auflösungsverhalten von nichtlinearen Fixmengen-Systemen, in *Iterative solution of nonlinear systems of equations* (Oberwolfach, 1982), vol. 953 of Lecture Notes in Math., Springer, Berlin, (1982), 106–124.

[35] D. Pallaschke and R. Urbański, *Pairs of Compact Convex Sets*, vol. 548 of Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 2002.

[36] G. Perria, *Set-valued Interpolation*, vol. 79 of Bayreuth. Math. Schr., Department of Mathematics, University of Bayreuth, Germany, 2007.

[37] B. T. Polyak, Convexity of quadratic transformations and its use in control and optimization, *J. Optim. Theory Appl.*, **99** (1998), 553–583.

[38] B. T. Polyak, Convexity of nonlinear image of a small ball with applications to optimization, *Set-Valued Anal.*, **9** (2001), 159–168, Special issue on Wellposedness in Optimization and Related Topics (Gargnano, 1999).

[39] L. S. Pontryagin, Linear differential games. II, *Sov. Math., Dokl.*, **8** (1967), 910–912.

[40] R. T. Rockafellar, *Convex Analysis*, vol. 28 of Princeton Mathematical Series, Princeton University Press, Princeton, NJ, 1970.

[41] A. M. Rubinov and I. S. Akhundov, Difference of compact sets in the sense of Demyanov and its application to non-smooth analysis, *Optimization*, **23** (1992), 179–188.

[42] P. Saint-Pierre, Newton and other continuation methods for multivalued inclusions, *Set-Valued Anal.*, **3** (1995), 143–156.

[43] C. I. Siettos, C. C. Pantelides and I. G. Kevrekidis, Enabling dynamic process simulators to perform alternative tasks: A time-stepper-based toolkit for computer-aided analysis, *Ind. Engng. Chem. Res.*, **42** (2003), 6795–6801.

[44] C. Theodoropoulos, Y.-H. Qian and I. G. Kevrekidis, “Coarse” stability and bifurcation analysis using time-steppers: A reaction-diffusion example, *Proc. Natl. Acad. Sci.*, **97** (2000), 9840–9843, [http://www.pnas.org/content/97/18/9840.abstract](http://www.pnas.org/content/97/18/9840.abstract).

[45] L. S. Tuckerman and D. Barkley, Bifurcation analysis for timesteppers, in *Numerical methods for bifurcation problems and large-scale dynamical systems* (Minneapolis, MN, 1997), vol. 119 of IMA Vol. Math. Appl., Springer, New York, (2000), 453–466.

[46] X. Wang, Convergence of Newton’s method and inverse function theorem in Banach space, *Math. Comp.*, **68** (1999), 169–186.

[47] T. Yamamoto, A method for finding sharp error bounds for Newton’s method under the Kantorovich assumptions, *Numer. Math.*, **49** (1986), 203–220.

Received July 2012; revised February 2014.

E-mail address: robert.baier@uni-bayreuth.de; dellnitz@uni-paderborn.de
E-mail address: mirkoh@uni-paderborn.de; sertl@uni-paderborn.de; yannis@princeton.edu
Figure 2. Successive iterates of three Newton iterations computing invariant sets for Example 6.1. Left: Case (a) Starting with an ellipse lying “far out”, the iteration converges to $B_2(0)$. Middle (case (c)) and right (case (b)): Iterations converging to $B_1(0)$.
Figure 3. Visualizations of the differences between the first two iterates in the iterations for example 6.1. Left: Case (a), cf. Fig. 2 left. Right: Case (c), cf. Fig. 2 middle.
(a) Example 6.2(i), $\alpha = 0.1$

(b) Example 6.2(i), $\alpha = -0.1$

(c) Example 6.2(ii)

(d)

Figure 4. (a)–(c) Convex sets defined by the Newton iterations for Example 6.2 converging to the convex invariant sets of the corresponding underlying dynamical systems. (d) Comparison of an approximation of the convex invariant set of the map $g$ of Example 6.2(ii) computed via the subdivision algorithm from [12] (blue area) and the boundary of the convex set computed via Newton’s method (black line).