A "Converse" of the Banach Contraction Mapping Theorem

Pascal Hitzler*

Artificial Intelligence Institute
Dresden University of Technology
Germany
phitzler@inf.tu-dresden.de
www.wv.inf.tu-dresden.de/~pascal/

Anthony Karel Seda†

Department of Mathematics
University College Cork
Ireland
a.seda@ucc.ie
maths.ucc.ie/staff/seda/

Abstract

We prove a type of converse of the Banach contraction mapping theorem for metric spaces: if X is a T_1 topological space and $f : X \to X$ is a function with unique fixed point a such that $f^n(x)$ converges to a for each $x \in X$, then there exists a distance function d on X such that f is a contraction on the complete ultrametric space (X, d) with contractivity factor $\frac{1}{2}$. We explore properties of the resulting space (X, d).

1 Introduction and Motivation

Fixed points of functions and of operators are important in many classical mathematical areas ranging from analysis to dynamical systems to geometry etc. One area in which they have recently risen to prominence is the area of denotational semantics of programming languages, or the process of giving precise mathematical meaning to computer programs, see for example [19] for theoretical background. To this end, semantic operators are associated with various constructs[1] in a program in such a way that the fixed points of these operators correspond to the denotational reading of the constructs. A particular case in point, and one which is the inspiration for the present work, is that of the semantics of logic programming systems, see [14]. In this context, an operator F can often be assigned to each program P in the system in such a manner that the

*The first named author acknowledges financial support under grant SC/98/621 from Enterprise Ireland.

†The second named author wishes to extend his thanks to the conference organizers, and to the Slovak University of Technology for the hospitality and support he received whilst attending SCAM2001.

[1]In particular, those relating to recursion and self-reference.

denotational or declarative semantics of the entire program P can be viewed in terms of the fixed points of F.

Of course, fixed-point theorems are of major importance anywhere that fixed points are important. In the theory of programming languages, the fixed-point theorems used, for example those of Knaster-Tarski and Kleene, are often based on order. Nevertheless, in recent years there has been a significant movement towards the use of theorems, such as the Banach contraction mapping theorem, which apply to metric spaces and especially to ultrametric spaces, and to variants and generalizations of ultrametric spaces, see for example [3, 4, 15]. This is particularly so in the case of logic programming, where the presence of negation may restrict the use of order theory. Fixed-point theorems on metric spaces and on generalized metric spaces provide an alternative and have indeed been applied successfully in the area of logic programming semantics, see [5, 7, 8, 12, 16, 17, 18]. In fact, investigations using tools from topology and analysis are of general interest in the area of logic programming, and elsewhere, in attempting to build continuous models of computation, see [2].

If the Banach contraction mapping theorem is applied in order to show that a function f has a unique fixed point a, then the proof of the Banach theorem also establishes the fact that iterates of f, or the sequences $(f^n(x))_n$, converge to a for each choice of x (see below for a statement of the Banach theorem). This strong property often allows one to compute approximations to the desired fixed point. In the context of logic programming, there are programs whose semantic operators are defined on topological spaces (of interpretations) and exhibit the same behaviour in that their iterates always converge (in the topological sense) to a unique fixed point, although no metric space structure has been given a priori and indeed may not be obviously available. In fact, on the contrary, we will show in this paper that such an ultrametric space underlying these processes can always be found. Thus, our main result, which is stated precisely in Theorem 2, is in a sense a converse of the Banach contraction mapping theorem, and permits that theorem to be "applied" in some circumstances where no metric rendering the operators in question to be contractions was readily to hand in advance.

It should be noted that our theorem is not a true converse of the Banach theorem in that we do not start out with a metrizable space and attempt to obtain a metric for it relative to which f is a contraction. Our results, therefore, are very different from those discussed in Section 3.6 of the text [13] which discusses a number of converses of the Banach theorem. Even the result of Bessaga discussed there, which applies to an abstract set, is very different from ours in that we do not require all iterations of f to have a unique fixed point, but we do require topological convergence of the iterates of any point. However, we wish to thank Professor Finbarr Holland for drawing our attention to the results of [13].

Our results may be particularly interesting in the context of relationships between logic programs and artificial neural networks [10, 12]: under certain conditions involving the requirement that the semantic operator satisfies the hypotheses of the Banach contraction mapping theorem, it can be shown that artificial neural networks exist which approximate this operator and its fixed point [12]. By providing a converse of the Banach contraction mapping theorem, we may be able to show that these hypotheses are satisfied in certain contexts, thus allowing us to apply the results from [12]. Such matters are under investigation by the authors.

The paper is structured as follows: we first recall some basic notions and establish some notation in the remainder of this section. In Section 2, we will state our main result and give its detailed proof. In Section 3, we will conclude with some further observations concerning the ultrametric space which results from Theorem 2.

For background on topological and metric spaces, and the Banach contraction mapping theorem, we refer to [20]. If X is a set, and $f : X \to X$ is a function, then, for all $x \in X$, we define inductively $f^0(x) = x$ and $f^{n+1}(x) = f(f^n(x))$ for $n \in \mathbb{N}$. Each $f^n(x)$ is called an *iterate* of f on x. We also define $f^{-1}(x) = \{y \in X \mid f(y) = x\}$ as usual. The set of all integers is denoted by $\mathbb{Z}$. Finally, we denote sequences by $(x_n)_{n\in\mathbb{N}}$ or simply by (x_n).

We recall the definition of a metric for later reference.

Definition 1 A *metric* on a set X is a function $d : X \times X \to \mathbb{R}$ which satisfies the following conditions (Mi) to (Miv) for all $x, y, z \in X$. We call d an *ultrametric* if it also satisfies (Mv), and we note that (Mv) implies (Miv).

(Mi) $d(x, y) = 0$ implies $x = y$.
(Mii) $d(x, x) = 0$.
(Miii) $d(x, y) = d(y, x)$.
(Miv) $d(x, z) \leq d(x, y) + d(y, z)$.
(Mv) $d(x, z) \leq \max\{d(x, y), d(y, z)\}$.

Finally, we recall the Banach contraction mapping theorem as follows.

Theorem 1 (Banach) Let (X, d) be a complete metric space and let $f : X \to X$ be a function such that there exists $0 \leq \lambda < 1$ with $d(f(x), f(y)) \leq \lambda d(x, y)$ for all $x, y \in X$. Then, for any $x \in X$, the sequence $(f^n(x))$ converges to the unique fixed point a of f.

Acknowledgement. The authors wish to thank an anonymous referee for drawing their attention to a gap in the first version of the proof of Theorem 2 which necessitated a slight change in the constructions made therein.

2 The Theorem

Theorem 2 Let (X, τ) be a T_1 topological space and let $f : X \to X$ be a function which has a unique fixed point a and is such that, for each $x \in X$, the sequence $(f^n(x))$ converges to a in τ. Then there exists a function $d : X \times X \to \mathbb{R}$ such that (X, d) is a complete ultrametric space and such that for all $x, y \in X$ we have $d(f(x), f(y)) \leq \frac{1}{2}d(x, y)$.

Proof: The proof is divided into several steps, numbered consecutively.

1. Given $x \in X$, we define the set $T(x) \subseteq X$ to be the smallest subset of X which is closed under the following rules.
(1) $x \in T(x)$.
(2) If $y \in T(x)$ and $f(y) \neq a$, then $f(y) \in T(x)$.
(3) If $y \in T(x)$ and $y \neq a$, then $f^{-1}(y) \subseteq T(x)$.
It is clear that the intersection of the family of all sets closed under these rules is itself closed under these rules, and so the minimality condition is met. Moreover, it is also clear that each of the sets $T(x)$ is non-empty. Now let $\mathcal{T} = \{T(x) \mid x \in X\}$ and observe the following facts:
(i) $T(a) = \{a\}$. To see this, we note that (1), (2) and (3) are all true relative to the set $\{a\}$. Therefore, by minimality we have $T(a) = \{a\}$.
(ii) If $x \neq a$, then $a \notin T(x)$ and so $T(a) \cap T(x) = \emptyset$. Hence, either $T(a)$ and $T(x)$ are equal or they are disjoint. To see this, suppose $x \neq a$ and consider rule (3). Clearly, we cannot have $a \in f^{-1}(x)$ otherwise $f(a) = x$ and hence $a = x$ which is a contradiction. Thus, rules (2) and (3) applied repeatedly and starting with x never place a in $T(x)$ and, by minimality, the process just described generates $T(x)$.
(iii) If $T(x) \neq T(a)$ and $T(y) \neq T(a)$, then either $T(x)$ and $T(y)$ are equal or they are disjoint. To see this, suppose $z \in T(x) \cap T(y)$. Then the rules (1), (2) and (3) under repeated application starting with z force $T(x) = T(y)$.
Thus, the collection $\mathcal{T}$ is a partition of X.

2. We next inductively define a mapping $l : T \to \mathbb{Z} \cup \{\infty\}$ on each $T \in \mathcal{T}$.
(1) We set $l(a) = \infty$, and this defines l on $T = T(a)$. If $T \neq T(a)$, we choose an arbitrary $x \in T$ and set $l(x) = 0$ (of course, $x \neq a$) and proceed as follows:
(2) For each $y \in T$ with $f(y) \neq a$ and $l(y) = k$, let $l(f(y)) = k + 1$.
(3) For each $y \in T$ with $l(y) = k$, let $l(z) = k - 1$ for all $z \in f^{-1}(y)$.
We will henceforth assume that all this is done for every $T \in \mathcal{T}$ so that l is a function defined on all of X. It is clear that the mapping l is well-defined since (X, τ) is a T_1 space[2]. For, if there is a cycle in the sequence $f^n(x)$ of iterates for some $x \in X$, then we can arrange for some element y in this sequence to be frequently not in some neighbourhood of a, using the fact that X is T_1, which contradicts the convergence of the sequence $f^n(x)$ to a.

3. Define a mapping $\iota : \mathbb{Z} \cup \{\infty\} \to \mathbb{R}$ by

$$\iota(k) = \begin{cases} 0 & \text{if } k = \infty, \\ 2^{-k} & \text{otherwise.} \end{cases}$$

Furthermore, define a mapping $\varrho : X \times X \to \mathbb{R}$ by

$$\varrho(x, y) = \max\{\iota(l(x)), \iota(l(y))\}$$

[2]We can weaken the requirement of τ being T_1 by replacing it with the following condition: for every $y \in X$ there exists an open neighbourhood U of a with $y \notin U$.

and a mapping $d : X \times X \to \mathbb{R}$ by

$$d(x,y) = \begin{cases} \varrho(x,y) & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

4. We show that (X,d) is an ultrametric space.

(Mi) Let $d(x,y) = 0$ and assume that $x \neq y$. Then $\varrho(x,y) = d(x,y) = 0$, hence $\max\{\iota(l(x)), \iota(l(y))\} = 0$, so $\iota(l(x)) = \iota(l(y)) = 0$. Hence $l(x) = l(y) = \infty$ and $x = y = a$ by construction of l, which is a contradiction.

(Mii) This is true by definition of d.

(Miii) This is true by symmetry of ϱ, and hence of d.

(Mv) Let $x,y,z \in X$ and assume without loss of generality that $\iota(l(x)) < \iota(l(z))$ so that $d(x,z) = \iota(l(z))$. If $\iota(l(y)) \leq \iota(l(z))$, then $d(y,z) = \iota(l(z))$. If $\iota(l(y)) > \iota(l(z))$, then $d(y,z) = \iota(l(y)) > \iota(l(z))$. In both cases we get $d(y,z) \geq d(x,z)$ as required.

5. (X,d) is complete as a metric space. In order to show this, let (x_n) be a Cauchy sequence in X. If (x_n) is eventually constant, then it converges trivially. So assume that (x_n) is not eventually constant. We show that x_n converges to a in d, for which it suffices to show that $(\iota(l(x_n)))_{n\in\mathbb{N}}$ converges to 0. Let $\varepsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ such that for all $m,n \geq n_0$ we have $d(x_m,x_n) < \varepsilon$. In particular, we have $d(x_m,x_{n_0}) < \varepsilon$ for all $m \geq n_0$, and, since (x_n) is not eventually constant, we obtain $\iota(l(x_{n_0})) < \varepsilon$, and also $\iota(l(x_m)) < \varepsilon$ for all $m \geq n_0$. Since ε was chosen arbitrarily, we see that $(\iota(l(x_n)))_{n\in\mathbb{N}}$ converges to 0.

6. We note that for $f(x) \neq a$, we have $l(f(x)) = l(x) + 1$ by definition of l, and hence $\iota(l(f(x))) = \frac{1}{2}\iota(l(x))$.

7. For all $x,y \in X$, we have that $d(f(x),f(y)) \leq \frac{1}{2}d(x,y)$. In order to show this, let $x,y \in X$ and assume without loss of generality that $x \neq y$. Let $d(x,y) = 2^{-k}$, say, so that $\max\{\iota(l(x)), \iota(l(y))\} = 2^{-k}$. Then $d(f(x),f(y)) = \max\{\iota(l(f(x))), \iota(l(f(y)))\} = \frac{1}{2}\max\{\iota(l(x)), \iota(l(y))\} = \frac{1}{2}d(x,y)$, as required. ■

3 Exploring the Proof

We make some observations concerning the proof of Theorem 2, and continue to use the notation adopted there.

i. Any $x \neq a$ is an isolated point with respect to d, that is, $\{x\}$ is open and closed in the topology generated by d.

Proof: Let $x \neq a$ and let $\iota(l(x)) = 2^{-k}$, say. Then, for any $y \in X$, we have $\varrho(x,y) \geq 2^{-k}$ and hence, for each $y \neq x$, we have $d(x,y) \geq 2^{-k}$. Therefore, $\{y \in X \mid d(x,y) < 2^{-k}\} = \{x\}$ which is consequently open in d. Closedness is trivial. ■

ii. If (x_n) is a sequence in X which converges in d to some $x \neq a$, then the sequence (x_n) is eventually constant.

Proof: In order to see this, it suffices to show that $\{x\}$ is open with respect to d for any $x \neq a$, which is true by point i. ■

iii. The metric d does not in general generate τ, but the iterates $(f^n(x))$ of f converge to a both with respect to τ and with respect to d.

Proof: Indeed, the topology τ is not in general metrizable. By the proof of the Banach contraction mapping theorem, $(f^n(x))$ converges to a with respect to d. Convergence with respect to τ follows from the hypothesis of Theorem 2. ■

iv. ϱ is a dislocated ultrametric.

Dislocated metrics, and the slightly stronger partial metrics, are a generalization of metrics and have recently been studied in the context of domain theory, see for example [6, 9, 15]. A *dislocated ultrametric* on a set X is a mapping $p : X \times X \to \mathbb{R}_0^+$, where $\mathbb{R}_0^+$ stands for the positive real numbers including zero, such that conditions (Mi), (Miii) and (Mv) from Definition 1 are satisfied. The assertion is easily derived from the definition of ϱ.

v. There exists a partial order $\leq$ on X such that the following hold:

(i) $(X, \leq)$ is chain complete, that is, every increasing chain has a least upper bound.
(ii) f is monotonic: if $x \leq y$, then $f(x) \leq f(y)$.
(iii) For each increasing chain (x_λ) in X, we have $f(\sup_\lambda x_\lambda) = \sup_\lambda f(x_\lambda)$.
(iv) $f(x) \geq x$ for all $x \in X$.

Proof: We define the partial order $\leq$ on X by (1) $x \leq a$ for all $x \in X$, and (2) $x \leq y$ if $l(x) \leq l(y)$ and $y \in T(x)$.

(i) Note that if $(x_n)_n$ is an increasing chain, then so is $(l(x_n))_n$. If $(x_n)_n$ is eventually constant, then there is nothing to show, so let (x_n) be strictly increasing. Then $(l(x_n))$ is also strictly increasing, and it is easy to see that a is the least upper bound of $(l(x_n))$.
(ii) If $l(x) \leq l(y)$, then $l(f(x)) = l(x) + 1 \leq l(y) + 1 = l(f(y))$.
(iii) If (x_λ) is eventually constant, then there is nothing to show. Otherwise, we have $f(\sup_\lambda(x_\lambda)) = f(a) = \sup_\lambda(f(x_\lambda))$, as required.
(iv) This is clear by definition of l.

Thus, the proof is complete. ■

We note the similarity between these conditions and the hypothesis of the Kleene theorem, see [1, 4].

vi. We finally note that Theorem 2 is essentially a generalization of a proposition in [11], which was proven for finite X.

References

[1] Baranga, A.: The Contraction Principle as a Particular Case of Kleene's Fixed-Point Theorem, Discrete Mathematics **98** (1991), 75–79.

[2] Blair, H.A., Dushin, F., Jakel, D.W., Rivera, A.J. and Sezgin, M.: Continuous Models of Computation for Logic Programs, In: Apt, K.R., Marek, V.W., Truszczyński, M. and Warren, D.S. (Eds.), The Logic Programming Paradigm: A 25-Year Perspective, Springer, Berlin, 1999, pp. 231–255.

[3] Bonsangue, M.M., van Breugel, F. and Rutten, J.J.M.M.: Alexandroff and Scott Topologies for Generalized Metric Spaces, In: Andima, S. et al. (Eds.), Papers on General Topology and Applications: Eleventh Summer Conference at University of Southern Maine, Annals of the New York Academy of Sciences, Vol. 806, 1996, pp. 49–68.

[4] Edalat, A. and Heckmann, R.: A Computational Model for Metric Spaces, Theoretical Computer Science **193** (1998), 53–73.

[5] Fitting, M.: Metric Methods: Three Examples and a Theorem, Journal of Logic Programming **21** (3) (1994), 113–127.

[6] Heckmann, R.: Approximation of Metric Spaces by Partial Metric Spaces, Applied Categorical Structures **7** (1999), 71–83.

[7] Hitzler, P.: Generalized Metrics and Topology in Logic Programming Semantics, Ph.D Thesis, National University of Ireland, University College Cork, 2001.

[8] Hitzler, P. and Seda, A.K.: The Fixed-Point Theorems of Priess-Crampe and Ribenboim in Logic Programming, In: Proceedings of the International Conference and Workshop on Valuation Theory, University of Saskatchewan in Saskatoon, Canada, July 1999. Fields Institute Communications Series, American Mathematical Society, Providence, R.I., 17 pages, to appear.

[9] Hitzler, P. and Seda, A.K.: Dislocated Topologies, In: Proceedings of SCAM 2000, Bratislava, April, 2000. Journal of Electrical Engineering 51 (12/s) (2000), pp. 3–7.

[10] Hitzler, P. and Seda, A.K.: A Note on the Relationships between Logic Programs and Neural Networks, In: Gibson, P. and Sinclair, D. (Eds.), Proceedings of the Fourth Irish Workshop on Formal Methods (IWFM'00), July 2000, Electronic Workshops in Computing (eWiC), British Computer Society, pp. 1-9, to appear.

[11] Hölldobler, S. and Kalinke, Y.: Towards a Massively Parallel Computational Model for Logic Programming, In: Proceedings of ECAI94 Workshop on Combining Symbolic and Connectionist Processing, ECCAI, 1994, pp. 68–77.

[12] Hölldobler, S., Störr, H. and Kalinke, Y.: Approximating the Semantics of Logic Programs by Recurrent Neural Networks, Applied Intelligence **11** (1999), 45–58.

[13] Istrăţescu, V.I.: Fixed Point Theory – An Introduction, Mathematics and its Applications, Volume 7, D. Reidel Publishing Company, Dordrecht, Holland, 1981.

[14] Lloyd, J.W.: Foundations of Logic Programming, Second Edition, Springer-Verlag, Berlin, 1988.

[15] Matthews, S.G.: Partial Metric Topology, In: Andima, S. et al. (Eds.), Proceedings of the Eighth Summer Conference on General Topology and its Applications, Annals of the New York Academy of Sciences, Vol. 728, 1994, pp. 183–197.

[16] Priess-Crampe, S. and Ribenboim, P.: Ultrametric Spaces and Logic Programming, Journal of Logic Programming **42** (2000), 59–70.

[17] Seda, A.K.: Quasi-metrics and the Semantics of Logic Programs, Fundamenta Informaticae **29** (1) (1997), 97–117.

[18] Seda, A.K. and Hitzler, P.: Strictly Level-Decreasing Logic Programs, In: Butterfield, A. and Flynn, S. (Eds.), Proceedings of the Second Irish Workshop on Formal Methods (IWFM'98), July 1999, Electronic Workshops in Computing (eWiC), British Computer Society, 1999, pp. 1-18.

[19] Stoltenberg-Hansen, V., Lindström, I. and Griffor, E.R.: Mathematical Theory of Domains, Cambridge Tracts in Theoretical Computer Science No. 22, Cambridge University Press, Cambridge, 1994.

[20] Willard, S.: General Topology, Addison-Wesley, Reading, MA, 1970.

Pascal Hitzler (Dipl.-Math.), born 1971 in Germany, finished his Diplom in Mathematics at the University of Tübingen in December 1998. Then he was a research assistant and PhD student of A.K. Seda in Cork and submitted his thesis in January 2001. He is now a scientific assistant at the Artificial Intelligence Institute of the Dresden University of Technology. Besides his current research in mathematical foundations of artificial intelligence, he is also actively involved in running enhancement programmes for mathematically skilled high-school students.

Anthony Karel Seda (Dr., C.Math, FIMA) is a Senior Lecturer in the Department of Mathematics, University College Cork, Ireland. His research interests in the past were in the area of mathematical analysis, but now are concerned with the mathematical foundations of computer science and, in particular, of computational logic and of artificial intelligence. He is a member of several learned societies including the New York Academy of Sciences and the European Association for Theoretical Computer Science (EATCS), and he edits the "News from Ireland" column of the Bulletin of the EATCS.