Identifying Fixed Points in Recurrent Neural Networks using Directional Fibers: Supplemental Material on Theoretical Results and Practical Aspects of Numerical Traversal

Technical Report CS-TR-5051

Garrett E. Katz[1*] James A. Reggia[1,2]

[1]Department of Computer Science, University of Maryland, College Park, 20742
[2]UMIACS, University of Maryland, College Park, 20742

December 2016

Abstract

Fixed points of recurrent neural networks can represent many things, including stored memories, solutions to optimization problems, and waypoints along non-fixed attractors. As such, they are relevant to a number of neurocomputational phenomena, ranging from low-level motor control and tool use to high-level problem solving and decision making. Therefore, global solution of the fixed point equations can improve our understanding and engineering of recurrent neural networks. While local solvers and statistical characterizations abound, we do not know of any method for efficiently and precisely locating all fixed points of an arbitrary network. To solve this problem we have proposed a novel strategy for global fixed point location, based on numerical traversal of mathematical objects we defined called directional fibers [2]. This report supplements our results in [2] by presenting certain technical aspects of our method in more depth.

Acknowledgements: Supported by ONR award N000141310597. Thanks to Drs. Howard Elman and Giovanni Forni for helpful discussions.

*Email: gkatz@cs.umd.edu

1 Introduction

In the source paper [2] that is supplemented by this report, we propose a new method for locating fixed points in recurrent neural networks (RNNs) with arbitrary connection weights. As shown in [2], the method is found to be competitive and complementary to an existing baseline approach, often finding a different and larger set of fixed points. As such, it constitutes an effective new tool for better understanding recurrent neural network dynamics, and improving our ability to engineer them so that they possess desired neurocomputational properties in any given application.

Our method is based on mathematical objects called *directional fibers*, which contain the fixed points, and which can be numerically traversed. Accurate and efficient traversal relies on a carefully derived numerical update scheme, and accurately counting the fixed points found requires special attention to round-off errors. This report covers these technical aspects.

For the reader's benefit, Section 2 recapitulates the relevant formal definitions from the source paper [2]. Section 3 derives the numerical update scheme for traversal along a directional fiber. Section 4 addresses the issue of round-off errors in finite-precision machine arithmetic.

2 Directional Fibers

In the following we assume the reader is already familiar with the terminology, concepts, and proofs given in [2]. For the reader's convenience we repeat the relevant definitions here without elaboration.

Definition 1. *Given a function* $f\colon \mathbb{R}^N \to \mathbb{R}^N$*, and* $c \in \mathbb{R}^N - \{\mathbf{0}\}$*, the* ***directional fiber*** *of* c *under* f *is defined as:*

$$\gamma^{(c)} \stackrel{\text{def}}{=} \{v \in \mathbb{R}^N \,:\, f(v) \text{ is parallel to } c\}, \tag{1}$$

or equivalently, as:

$$\Gamma^{(c)} \stackrel{\text{def}}{=} \{(v,\alpha) \in \mathbb{R}^N \times \mathbb{R} \,:\, F^{(c)}(v,\alpha) = \mathbf{0}\}, \tag{2}$$

where the function $F^{(c)} : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ *is given by:*

$$F^{(c)}(v,\alpha) \stackrel{\text{def}}{=} f(v) - \alpha c. \tag{3}$$

Where convenient, we rewrite points $(v,\alpha) \in \mathbb{R}^N \times \mathbb{R}$ as points $x \in \mathbb{R}^{N+1}$. As shown in [2], directional fibers can be numerically traversed as long as c satisfies certain regularity conditions. For the remainder we focus on the RNN model with update rule

$$\Delta v = f(v) \stackrel{\text{def}}{=} \sigma(Wv) - v, \tag{4}$$

where $v \in \mathbb{R}^N$ is the vector of neural activations, $W \in \mathbb{R}^{N\times N}$ is the matrix of synaptic connection weights, and σ is coordinate-wise hyperbolic tangent, as in [2]. Note that f (and hence $F^{(c)}$, for every c) is differentiable.

3 Numerical Update Scheme

At a given point $x^{(0)} \in \Gamma^{(c)}$, let z denote the tangent vector to $\Gamma^{(c)}$. As shown in the source paper [2], if the Jacobian $DF^{(c)}(x^{(0)})$ is full rank, then z is the unique (up to sign) unit vector satisfying

$$DF^{(c)}(x^{(0)})z = \mathbf{0}. \tag{5}$$

The numerical step advances $x^{(0)}$ by a distance of θ^* in the direction of z, resulting in a new point $x^{(\theta^*)} \in \Gamma^{(c)}$. Our update scheme accomplishes this by using Newton's method to solve

$$G(x^{(\theta^*)}) = \begin{bmatrix} \mathbf{0} \\ \theta^* \end{bmatrix} \tag{6}$$

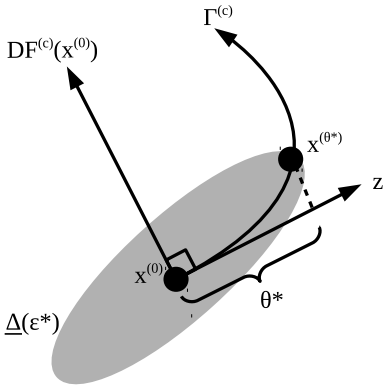

Figure 1: An illustration of some key quantities used in Theorem 1. The arrow labeled "$DF^{(c)}(x^{(0)})$" represents the row space of $DF^{(c)}$, which is orthogonal to z (in higher dimensions there will be more than one row vector).

for $x^{(\theta^*)}$, seeded with $x^{(0)}$, where $G : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ is defined by

$$G(x) \stackrel{\text{def}}{=} \begin{bmatrix} F^{(c)}(x) \\ z^T(x - x^{(0)}) \end{bmatrix}. \tag{7}$$

Eq. 6 simultaneously maintains $F^{(c)}(x^{(\theta^*)}) = \mathbf{0}$, which keeps $x^{(\theta^*)}$ in $\Gamma^{(c)}$, and enforces $z^T(x^{(\theta^*)} - x^{(0)}) = \theta^*$, which moves the traversal forward by a distance of θ^* in the tangent direction. This update is illustrated in Fig. 1.

As long as W is invertible, the step-size θ^* can be determined rigorously with strong formal guarantees. In particular, this section shows how to compute a θ^* for which the numerical update is *guaranteed to converge to the same point that would have resulted from the mathematically ideal traversal*: that is, the traversal in which $x^{(0)}$ flows continously along $\Gamma^{(c)}$, by a distance of θ^*, in the direction of z. The conditions that θ^* must satisfy for this to hold are provided by Theorem 1 below. For greater notational ease in the statement and proof of this theorem, we define several auxiliary functions and quantities as follows, some of which are shown in Fig. 1.

First we let λ denote the smallest singular value of $DG(x^{(0)})\tilde{W}^{-1}$, where DG is the Jacobian of G and $\tilde{W}$ abbreviates

$$\begin{bmatrix} W & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{8}$$

Next, given any $\varepsilon > 0$, we define $\delta_i(\varepsilon) > 0$ to be the largest δ such that for $i \leq N$ and any $x \in \mathbb{R}^{N+1}$, if

$$|\tilde{W}_{i,:}(x - x^{(0)})| < \delta, \tag{9}$$

then

$$|\sigma'(\tilde{W}_{i,:}x) - \sigma'(\tilde{W}_{i,:}x^{(0)})| < \varepsilon. \tag{10}$$

$\delta_i(\varepsilon)$ is used to determine a neighborhood around $x^{(0)}$ in which $DG(x)$ remains close to $DG(x^{(0)})$, where "closeness" is measured by ε. Its computation is explained after the statement of the theorem and illustrated in Fig. 2.

Based on $\delta_i(\varepsilon)$, we define several intermediate bounds used by the theorem:

- $\Delta_i(\varepsilon) \stackrel{\text{def}}{=} \{x : |\tilde{W}_{i,:}(x - x^{(0)})| < \delta_i(\varepsilon)\}$,
- $\underline{\delta}(\varepsilon) \stackrel{\text{def}}{=} \min_i \delta_i(\varepsilon)$,

- $\underline{\Delta}(\varepsilon) \stackrel{\text{def}}{=} \{x : ||\tilde{W}(x - x^{(0)})|| < \underline{\delta}(\varepsilon)\}$,
- $\mu(\varepsilon) \stackrel{\text{def}}{=} \max_i \max_{x \in \Delta_i(\varepsilon)} \frac{1}{2}|\sigma''(\tilde{W}_{i,:}x)|$,
- and $\rho(\varepsilon) \stackrel{\text{def}}{=} \mu(\varepsilon)/(\lambda - \varepsilon)$.

Note that $\delta_i, \underline{\delta}, \mu$, and ρ are all positive for $\varepsilon \in (0, \lambda)$.

Finally, we define

$$\Theta(\varepsilon) \stackrel{\text{def}}{=} \frac{1}{||\tilde{W}z||} \cdot \frac{\underline{\delta}(\varepsilon)}{1 + \rho(\varepsilon)\underline{\delta}(\varepsilon)}, \tag{11}$$

we let

$$\varepsilon^* = \operatorname*{argmax}_{\varepsilon \in (0,\lambda)} \Theta(\varepsilon), \tag{12}$$

and we let $\theta^* = \Theta(\varepsilon^*)$.

Theorem 1. *Given fixed c and invertible W, let $x^{(0)}$ be any point in $\Gamma^{(c)}$. Suppose $DF^{(c)}(x^{(0)})$ is full rank and z is the tangent vector spanning its null space. Then for each $\theta \in [0, \theta^*]$, there is a unique $x^{(\theta)} \in \underline{\Delta}(\varepsilon^*)$ satisfying*

$$G(x^{(\theta)}) = \begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix}, \tag{13}$$

and Newton's method, when seeded with $x^{(0)}$ and used to solve Eq. (13), will converge to $x^{(\theta)}$. Moreover, the resulting bijection $\theta \mapsto x^{(\theta)}$ is continuous on $[0, \theta^*]$.

In Theorem 1, each $x^{(\theta)}$ solving Eq. (13) is a point in $\Gamma^{(c)}$ that lies a distance of θ from $x^{(0)}$ in the direction of the tangent z. The fact that the map $\theta \mapsto x^{(\theta)}$ is a continuous bijection for all $\theta \in [0, \theta^*]$ guarantees that the same $x^{(\theta)}$ would result from the mathematically ideal traversal where x flows continuously along $\Gamma^{(c)}$ starting from $x^{(0)}$. The proof follows a common strategy of proving the IVT, based on Newton's method (e.g., [3, 4]). However we take additional care to keep an explicit bound on the region of convergence as large as possible, capitalizing on the specific characteristics of the network model studied here. In this proof the n^{th} iterate of Newton's method is denoted $x^{(n)}$. Whereas we solve for $x^{(\theta^*)}$ during fiber traversal since it is the largest step-size with a formal guarantee, in this proof we solve for $x^{(\theta)}$, for an arbitrary $\theta \in [0, \theta^*]$, to establish the guarantee.

Proof of Theorem 1. Let $\rho^*, \mu^*, \underline{\delta}^*, \underline{\Delta}^*$ abbreviate $\rho(\varepsilon^*), \mu(\varepsilon^*), \underline{\delta}(\varepsilon^*), \underline{\Delta}(\varepsilon^*)$. By rearranging Eq. (11), we get both

$$\rho^*||\tilde{W}z||\theta^* = \frac{\underline{\delta}^*}{\underline{\delta}^* + 1/\rho^*} < 1 \tag{14}$$

and

$$\frac{\rho^*||\tilde{W}z||\theta^*}{1 - \rho^*||\tilde{W}z||\theta^*} = \rho^*\underline{\delta}^*. \tag{15}$$

Now consider any $\theta \in [0, \theta^*]$. Given Eq. (14), the left-hand side of Eq. (15) is the closed form for a geometric series with ratio $\rho^*||\tilde{W}z||\theta^*$. Combining with $\theta \leq \theta^*$, we have

$$\frac{1}{\rho^*} \sum_{k=1}^{\infty} (\rho^*||\tilde{W}z||\theta)^k \leq \underline{\delta}^*. \tag{16}$$

Since $r^{2^k} \leq r^{k+1}$ for any positive r less than 1 and integer $k \geq 0$, Eq. (16) implies

$$\frac{1}{\rho^*} \sum_{k=0}^{\infty} (\rho^*||\tilde{W}z||\theta)^{2^k} \leq \underline{\delta}^*. \tag{17}$$

We will bound the Newton iterates within $\underline{\Delta}^*$ using Eq. (17) as well as the following bound on the derivatives of G. Let x be any point in $\underline{\Delta}^*$. Explicitly differentiating G, we have

$$DG(x) = \begin{bmatrix} \Sigma'(x)W - I, & -c \\ z^T & \end{bmatrix}, \tag{18}$$

where $\Sigma'(x)$ abbreviates $\text{diag}_{i \leq N}(\sigma'(\tilde{W}_{i,:}x))$. By adding and subtracting $DG(x^{(0)})\tilde{W}^{-1}$, we have

$$DG(x)\tilde{W}^{-1} = DG(x^{(0)})\tilde{W}^{-1} + \begin{bmatrix} (\Sigma'(x) - \Sigma'(x^{(0)})) & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}. \tag{19}$$

Since $x \in \underline{\Delta}^*$, we have

$$|\tilde{W}_{i,:}(x - x^{(0)})| \leq ||\tilde{W}(x - x^{(0)})|| \leq \underline{\delta}^* \leq \delta_i(\varepsilon^*) \tag{20}$$

for all $i \leq N$, which implies

$$\max_{i \leq N} |\sigma'(\tilde{W}_{i,:}x) - \sigma'(\tilde{W}_{i,:}x^{(0)})| \leq \varepsilon^* < \lambda \tag{21}$$

by the definition of δ_i and the constraint in Eq. (12) that $\varepsilon^* \in (0, \lambda)$. Therefore

$$\left|\left|\begin{bmatrix} (\Sigma'(x) - \Sigma'(x^{(0)})) & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}\right|\right| < \varepsilon^* < \lambda, \tag{22}$$

and combining with Eq. (19), we get

$$s^* > \lambda - \varepsilon^*, \tag{23}$$

where s^* is the minimal singular value of $DG(x)\tilde{W}^{-1}$.

We are now prepared to show that the Newton iterates converge. We will prove by induction that

$$||\tilde{W}(x^{(n+1)} - x^{(n)})|| \leq (\rho^*||\tilde{W}z||\theta)^{2^n}/\rho^* \tag{24}$$

for all iterates $x^{(n)}$. The induction relies on the formula for Newton iterations, which can be expressed as

$$\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} - G(x^{(n)}) = DG(x^{(n)})(x^{(n+1)} - x^{(n)}). \tag{25}$$

Eq. (25) is solved for $x^{(n+1)}$ on each iteration.

In the base case $n = 0$, writing Eq. (25) more explicitly, we have:

$$\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} = \begin{bmatrix} DF^{(c)}(x^{(0)}) \\ z^T \end{bmatrix}(x^{(1)} - x^{(0)}), \tag{26}$$

which is solved by $x^{(1)} - x^{(0)} = z\theta$, since z is a unit vector spanning the null space of $DF^{(c)}(x^{(0)})$. Therefore

$$||\tilde{W}(x^{(1)} - x^{(0)})|| = ||\tilde{W}z||\theta = (\rho^*||\tilde{W}z||\theta)^{2^0}/\rho^*, \tag{27}$$

and Eq. (24) is true with equality.

For the inductive case, suppose Eq. (24) is true for $k \leq n$. Then we have

$$||\tilde{W}(x^{(k)} - x^{(0)})|| \leq \sum_{j=0}^{k-1} ||\tilde{W}(x^{(j+1)} - x^{(j)})|| \tag{28}$$

$$\leq \sum_{j=0}^{k-1} (\rho^*||\tilde{W}z||\theta)^{2^j}/\rho^* \tag{29}$$

$$\leq \underline{\delta}^* \tag{30}$$

where Eqs. (28-30) follow by the triangle inequality, the inductive hypothesis, and Eq. (17), respectively. This shows that $x^{(n)}$ and $x^{(n-1)}$ are both in $\underline{\Delta}^*$.

Using $x^{(n)}, x^{(n-1)} \in \underline{\Delta}^*$ we derive a recursive relation on the iterates as follows. Recapitulating Eq. (25), the n^{th} and $(n+1)^{th}$ Newton iterates are computed according to

$$\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} - G(x^{(n-1)}) = DG(x^{(n-1)})(x^{(n)} - x^{(n-1)}) \tag{31}$$

$$\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} - G(x^{(n)}) = DG(x^{(n)})(x^{(n+1)} - x^{(n)}). \tag{32}$$

Subtracting (32) from (31) gives

$$G(x^{(n)}) - G(x^{(n-1)}) = DG(x^{(n-1)})(x^{(n)} - x^{(n-1)}) - DG(x^{(n)})(x^{(n+1)} - x^{(n)}). \tag{33}$$

By Taylor's theorem [1], $G(x^{(n)}) - G(x^{(n-1)})$ also satisfies

$$G(x^{(n)}) - G(x^{(n-1)}) = DG(x^{(n-1)})(x^{(n)} - x^{(n-1)}) + R^{(n-1)}, \tag{34}$$

with second-order remainder term $R^{(n-1)}$. Substituting (33) into (34) and canceling terms leaves

$$-DG(x^{(n)})(x^{(n+1)} - x^{(n)}) = R^{(n-1)}, \tag{35}$$

and explicitly differentiating DG shows that for $i \leq N$,

$$R_i^{(n-1)} = \frac{1}{2}\sigma''(\tilde{W}_{i,:}\tilde{x}^{(i,n)})(\tilde{W}_{i,:}(x^{(n)} - x^{(n-1)}))^2, \tag{36}$$

where each $\tilde{x}^{(i,n)}$ is a weighted average of $x^{(n)}$ and $x^{(n-1)}$, and hence also in $\underline{\Delta}^*$. As for $i = N+1$, differentiation shows that $R_{N+1}^{(n-1)} = 0$.

Inserting the product $\tilde{W}^{-1}\tilde{W} = I$ in the left-hand side of Eq. (35) and taking the norm of both sides, we have

$$||DG(x^{(n)})\tilde{W}^{-1}\tilde{W}(x^{(n+1)} - x^{(n)})|| = ||R^{(n-1)}||. \tag{37}$$

From Eq. (23), this implies

$$(\lambda - \varepsilon^*)||\tilde{W}(x^{(n+1)} - x^{(n)})|| \leq ||R^{(n-1)}||. \tag{38}$$

To bound $||R^{(n-1)}||$, we first note that since each $\tilde{x}^{(i,n)} \in \underline{\Delta}^*$, we have $\max_i \frac{1}{2}\sigma''(\tilde{W}_{i,:}\tilde{x}^{(i,n)}) \leq \mu^*$. Moreover, for any vector a, we have $||a||^2 \geq ||a^2||$, where the exponent inside the norm is taken coordinate-wise. This is true because

$$(||a||^2)^2 = \left(\sum_i a_i^2\right)^2 = \sum_i a_i^4 + \sum_{i \neq j} a_i^2 a_j^2 \geq \sum_i a_i^4 = ||a^2||^2. \tag{39}$$

Finally, $||R^{(n-1)}|| = ||R_{1:N}^{(n-1)}||$ since $R_{N+1}^{(n-1)} = 0$. Therefore from Eqs. (36), (38), and (39), we get

$$||\tilde{W}(x^{(n+1)} - x^{(n)})|| \leq \frac{\mu^*}{\lambda - \varepsilon^*}||\tilde{W}(x^{(n)} - x^{(n-1)})||^2 = \rho^*||\tilde{W}(x^{(n)} - x^{(n-1)})||^2. \tag{40}$$

Substituting from the inductive hypothesis on the right-hand side of Eq. (40), we have

$$||\tilde{W}(x^{(n+1)} - x^{(n)})|| \leq \rho^* \left((\rho^*||\tilde{W}z||\theta)^{2^{n-1}})/\rho^*\right)^2 = (\rho^*||\tilde{W}z||\theta)^{2^n}/\rho^*. \tag{41}$$

Hence the induction goes through for all n. The consequence is that $x^{(n)}$ is a Cauchy sequence, and therefore converges to a limit. It remains to show that the limit of $x^{(n)}$ is in fact a solution $x^{(\theta)}$ of

$$G(x^{(\theta)}) = \begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} \tag{42}$$

that is unique in $\underline{\Delta}^*$, and that the associated map $\theta \mapsto x^{(\theta)}$ is continuous.

To show that $\lim_{n\to\infty} x^{(n)}$ is a solution of Eq. (42), we again take norms in the Newton iteration formula, which gives

$$\left|\left|\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} - G(x^{(n)})\right|\right| \leq ||DG(x^{(n)})\tilde{W}^{-1}|| \cdot ||\tilde{W}(x^{(n+1)} - x^{(n)})||. \tag{43}$$

Since $||DG(x^{(n)})\tilde{W}^{-1}|| > \lambda - \varepsilon^* > 0$, whereas $||\tilde{W}(x^{(n+1)} - x^{(n)})||$ approaches 0, it must be that

$$\left|\left|\begin{bmatrix} \mathbf{0} \\ \theta \end{bmatrix} - G(x^{(n)})\right|\right| \tag{44}$$

also approaches 0 (and hence $x^{(n)}$ approaches a solution $x^{(\theta)}$).

For uniqueness, we take norms using the first-order Taylor theorem, which shows for any x that

$$||G(x^{(\theta)}) - G(x)|| = ||DG(\tilde{x})\tilde{W}^{-1}\tilde{W}(x^{(\theta)} - x)|| \tag{45}$$

$$\geq (\lambda - \varepsilon^*)||\tilde{W}(x^{(\theta)} - x)||, \tag{46}$$

where $\tilde{x}$ is a weighted average of $x^{(\theta)}$ and x and hence in $\underline{\Delta}^*$. Therefore if $||G(x^{(\theta)}) - G(x)|| = 0$, then it must be that $||\tilde{W}(x^{(\theta)} - x)|| = 0$. In other words, if $G(x^{(\theta)}) = G(x)$, then $x^{(\theta)} = x$.

Lastly, take any $e > 0$. To show continuity, we must find some $d > 0$, such that for any $\hat{\theta} \in [0, \theta^*]$,

$$|\theta - \hat{\theta}| < d \text{ implies } ||W(x^{(\theta)} - x^{(\hat{\theta})})|| < e. \tag{47}$$

Taking $x = x^{(\hat{\theta})}$ in Eq. (46), we obtain

$$||G(x^{(\theta)}) - G(x^{(\hat{\theta})})|| \geq (\lambda - \varepsilon^*)||\tilde{W}(x^{(\theta)} - x^{(\hat{\theta})}||. \tag{48}$$

Noting that $||G(x^{(\theta)}) - G(x^{(\hat{\theta})})|| = |\theta - \hat{\theta}|$, we find that setting $d = e(\lambda - \varepsilon^*)$ is sufficient. □

The quantities $\delta_i(\varepsilon)$, $\mu(\varepsilon)$, and $\rho(\varepsilon)$ can all be computed for any given ε with elementary, albeit cumbersome, operations, based on the properties of σ. Since $\sigma'(r) = 1 - \sigma^2(r)$ for any $r \in \mathbb{R}$, σ' can be inverted as follows:

$$r = (\sigma')^{-1}(\sigma'(r)) = \pm\sigma^{-1}\left(\sqrt{1 - \sigma'(r)}\right). \tag{49}$$

Using Eq. (49), $\delta_i(\varepsilon)$ can be computed as

$$\delta_i(\varepsilon) = \min\left\{\left|\pm\sigma^{-1}\left(\sqrt{1 - (\sigma'(\tilde{W}_i x^{(0)}) \pm \varepsilon)}\right) - \tilde{W}_i x^{(0)}\right|, \infty\right\}, \tag{50}$$

where the minimum is taken over all choices of $\pm$ that produce real-valued results (e.g., the horizontal lines in Fig. 2 that intersect the graph of σ'.). If none of the choices do, then $\delta_i(\varepsilon) = \infty$ signifies that any δ_i, no matter how large, satisfies the definition of $\delta_i(\varepsilon)$ in Theorem 1. Two examples of this computation are illustrated in Fig. 2.

Differentiation shows that we can compute $\sigma''(r)$ directly as $\sigma''(r) = 2\sigma'(r)\sigma(r) = 2(1 - \sigma^2(r))\sigma(r)$. Moreover, the maximum of $|\sigma''(r)|$ over any interval either occurs at one of the endpoints, or else is the global maximum of $|\sigma''(r)|$, namely $\sqrt{16/27}$, which occurs at $r = \sigma^{-1}(\sqrt{1/3})$. So $\mu(\varepsilon)$ can be computed as

$$\mu(\varepsilon) = \frac{1}{2}\begin{cases} \sqrt{16/27} & \text{if } \sigma^{-1}(\sqrt{1/3}) \in \Delta_i(\varepsilon) \\ \max_i \left|\sigma''(\tilde{W}_i x^{(0)} \pm \delta_i(\varepsilon))\right| & \text{otherwise,} \end{cases} \tag{51}$$

where the $\max_i$ is taken over each choice of sign for each i. This computation is illustrated in Fig. 3.

Once each δ_i and μ are computed, ρ can be computed directly from its definition. As for ε^*, it can be approximated reasonably well by evaluating Eq. (12) at a modest number (we used 16) of regularly spaced values of $\varepsilon \in (0, \lambda)$, thereby efficiently computing a step-size θ reasonably close to θ^*.

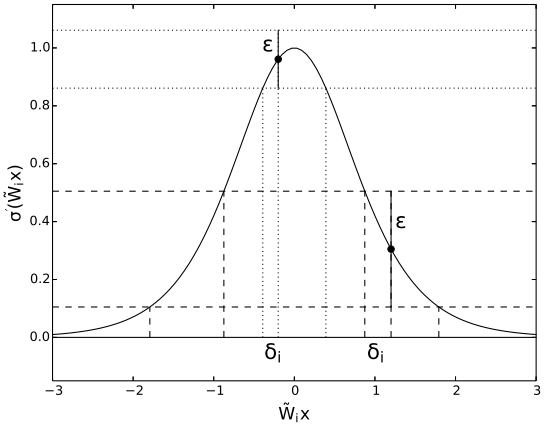

Figure 2: Two examples of computing δ_i from ε, one in dashed lines and one in dotted lines. First, $\sigma'(\tilde{W}_i x^{(0)}) \pm \varepsilon$ is calculated (horizontal lines), representing endpoints of the range in which we want to bound σ'. Then, the calculated endpoints are passed through $(\sigma')^{-1}$ (vertical lines), to obtain the endpoints of the corresponding range in which we should bound $\tilde{W}_i x$. These endpoints are subtracted from $\tilde{W}_i x^{(0)}$ to obtain δ_i.

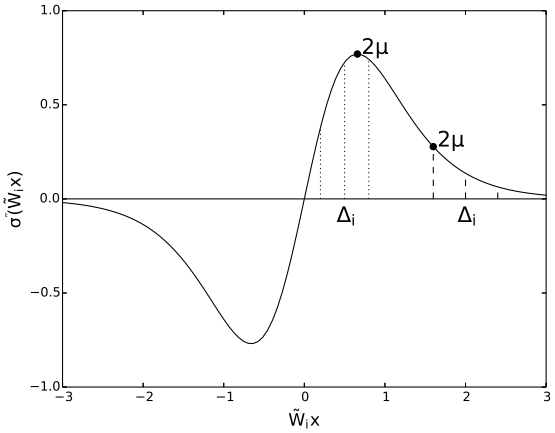

Figure 3: Two examples of computing μ from $\delta_i(\varepsilon)$, one in dashed lines and one in dotted lines. First, $\tilde{W}_i x^{(0)} \pm \delta_i$ is calculated to obtain the endpoints of Δ_i. Then each endpoint is passed through σ'' to determine μ (vertical lines). 2μ is either the greater of the two endpoints, or the global maximum of σ'' if it is included in Δ_i.

A corollary of Theorem 1 is that, while confined to $\underline{\Delta}(\varepsilon^*)$, the directional fiber cannot "double back" in the direction of $-z$ (otherwise, there would be two distinct $x^{(\theta)} \in \underline{\Delta}(\varepsilon^*)$ for the same θ, contradicting the theorem). In other words, the new tangent vector after the step should have a positive dot product with the previous tangent vector before the step. This allows us to ensure that the numerical traversal never inadvertently reverses direction from one step to the next. Specifically, we compute the new tangent vector after the step, which we denote $\hat{z}$, by solving the linear system

$$\begin{bmatrix} DF(x^{(\theta)}) \\ z^T \end{bmatrix} \hat{z} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \tag{52}$$

for $\hat{z}$ and then normalizing $\hat{z}$ to unit magnitude. This ensures both that $\hat{z}$ spans the null space of $DF(x^{(\theta)})$, so that it is tangent to $\Gamma^{(c)}$ at $x^{(\theta)}$, and also that $z^T \hat{z} > 0$, so that traversal continues in the right direction.

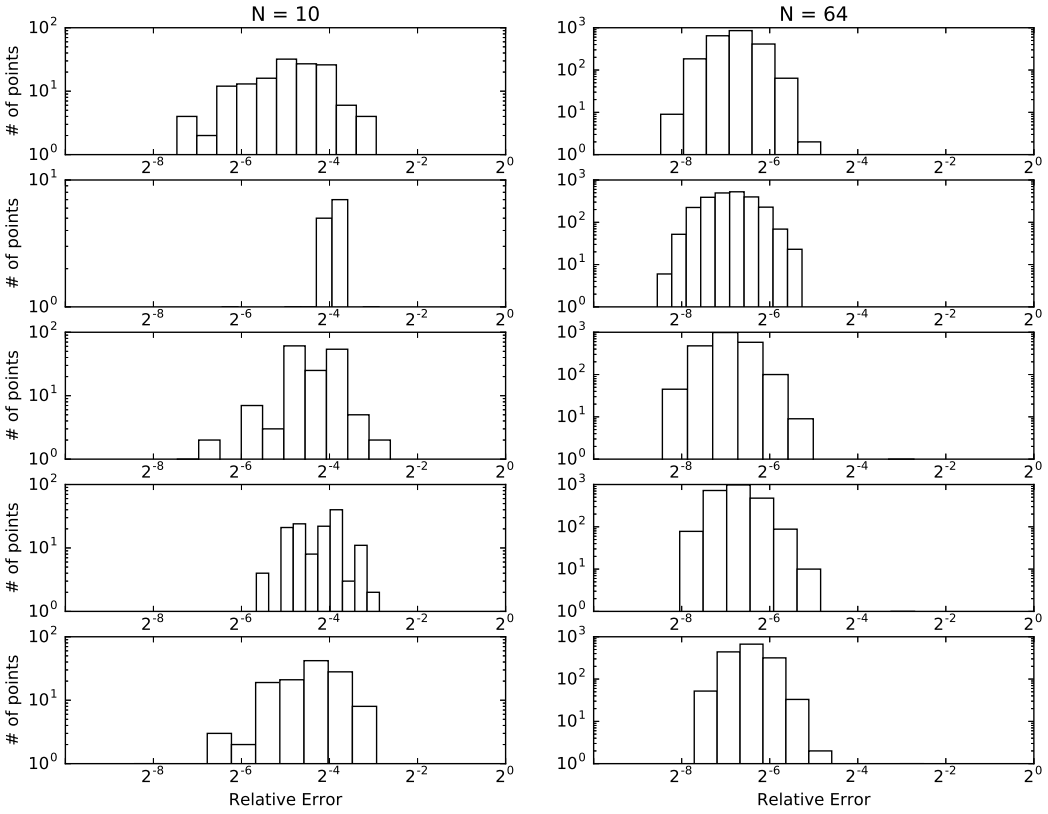

Figure 4: Histograms of relative errors $\mathcal{RE}(\overline{v})$ at points found by fiber traversal. Each and every fixed point was accepted as fixed by a large margin.

4 Counting Unique Fixed Points in Finite Precision

As described in the source paper [2], in order to accurately assess the performance of our solver, it is important to accurately count the number of unique fixed points found. Determining whether a point should be considered fixed, and whether two fixed points should be considered identical or distinct, are non-trivial problems in finite-precision arithmetic. The computed value of $f(v)$ at "fixed points" was generally a few multiples of machine precision and rarely identically $\mathbf{0}$. Similarly, any pair of "identical" fixed points were generally a few multiples of machine precision apart, and rarely identically equal.

To decide whether a point should be considered fixed, we use a forward error analysis of f to obtain the following upper bound:

$$|\overline{f(\overline{v})} - f(v)| \leq \mathcal{E}(\overline{v}) \stackrel{\text{def}}{=} \overline{|W|\epsilon(\overline{v})} + N\epsilon(\overline{|W||\overline{v}|}) + 5\epsilon(\overline{\sigma(\overline{W\overline{v}})}) + \epsilon(\overline{v}) + \max(\epsilon(\overline{\sigma(\overline{W\overline{v}})}), \epsilon(\overline{v})), \tag{53}$$

where all operations (except matrix multiplication) are applied coordinate-wise, and the inequality is true in every coordinate. The overbars denote the closest finite-precision approximation to an infinite-precision value, and $\epsilon(\cdot)$ denotes machine precision at a given finite-precision value. The coefficient of 5 bounds the relative error of σ. Rather than inspecting the machine implementation of hyperbolic tangent, we estimated this coefficient empirically based on the evaluation of $\sigma(\overline{x})$ at 2^{16} values of $\overline{x}$ uniformly sampled from $[0, 1]$. At a true fixed point v, $f(v) = \mathbf{0}$, and $|\overline{f(\overline{v})} - f(v)| = |\overline{f(\overline{v})} - \mathbf{0}| = |\overline{f(\overline{v})}|$, so any finite-precision point $\overline{v}$ satisfying $|\overline{f(\overline{v})}| > \mathcal{E}(v)$ can be rejected as certainly not fixed.

As a sanity check, we inspected histograms of the relative errors at points accepted and rejected as fixed

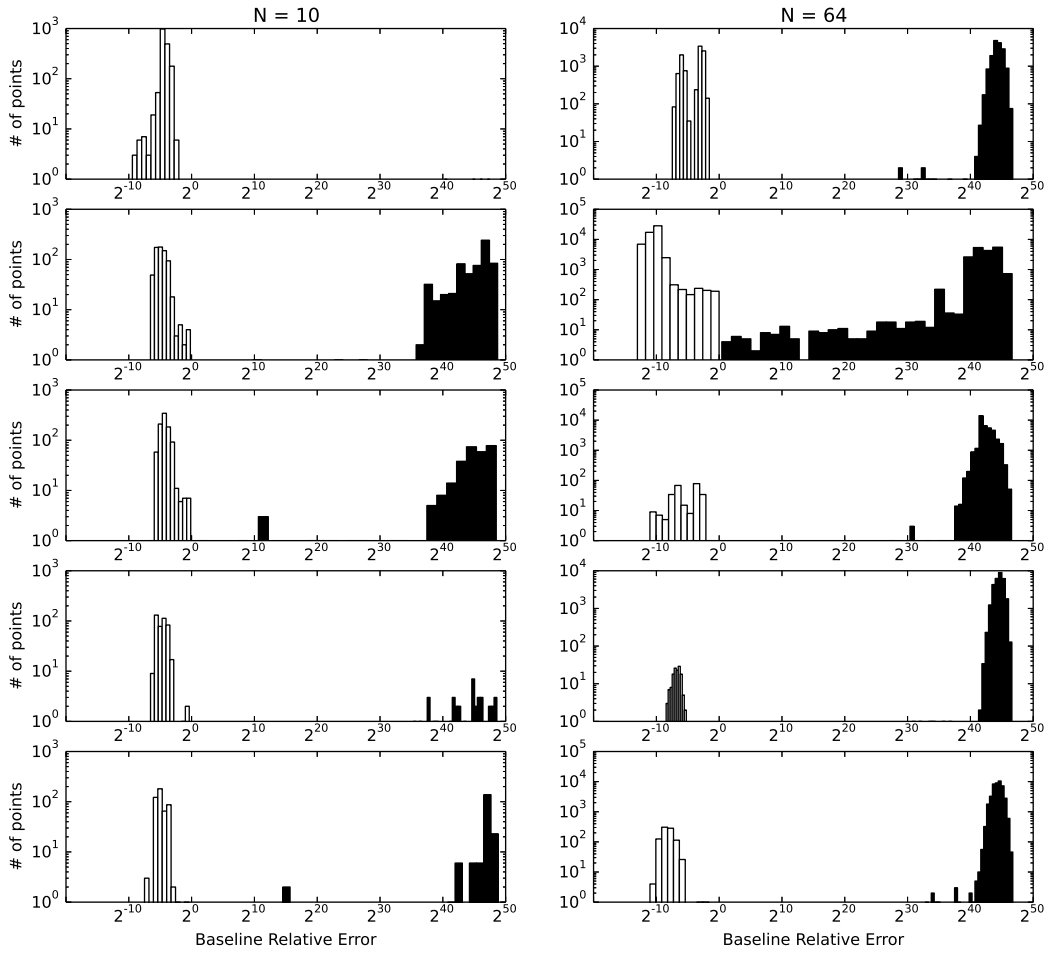

Figure 5: Histograms of relative errors $\mathcal{RE}(\overline{v})$ at points found by a baseline fixed point solver, colored according to whether they were accepted as fixed (white bars) or rejected as not fixed (black bars).

according to this test. We define the relative error $\mathcal{RE}$ as

$$\mathcal{RE}(\overline{v}) \stackrel{\text{def}}{=} \max_i |\overline{f_i(\overline{v})}|/\mathcal{E}_i(\overline{v}), \tag{54}$$

where the index i ranges over the coordinates of f (from 1 to N). Our test rejects $\overline{v}$ as certainly not within machine precision of a true fixed point if $\mathcal{RE}(\overline{v}) > 1$. Otherwise it accepts $\overline{v}$ as potentially within machine precision of a true fixed point. $\mathcal{RE}$ can be extremely large since $\mathcal{E}(\overline{v})$ is generally near machine precision, but when v is not fixed $f(v)$ can be much larger than machine precision.

Each panel in Fig. 4 shows the relative errors at "fixed" points found by a single fiber traversal on a single network. The left column contains panels for five networks with $N = 10$ and the right column contains panels for five networks with $N = 64$. The histograms show that each and every point identified by fiber traversal was accepted as fixed, by a wide margin, demonstrating that our theoretical results and error analysis are highly consistent.

Each panel in Fig. 5 shows the results for an existing baseline solver (described in [2]) on a single network. As in Fig. 4, the left column shows five networks with $N = 10$ and the right shows five networks with $N = 64$.

The baseline solver can locate either fixed points or so-called "slow" points that are not fixed but are local minima of $||f(v)||$. Points accepted as fixed are shown in white and points rejected as not fixed are shown in black. Although the distinction was typically clear cut, there were some edge cases which call the fidelity of our error analysis into question (middle panel on left, second panel from top on right). However, it is important to note that these histograms are shown with a log-scale on the y-axis. When shown normally, the edge cases are mostly invisible.

Nevertheless, some of our results in [2] rely on an accurate comparison of fiber traversal with the baseline solver. In particular, the results therein rely on the metrics $|T - B|$ and $|B - T|$, where T is the set of fixed points found by traversal and B is the set of points found by the baseline. $|T - B|$ measures the number of points that were found by the former but not the latter, and vice-versa for $|B - T|$. The edge cases in these histograms raise the question of whether allegedly larger values of $|T - B|$ than $|B - T|$ are in fact artifacts of a flawed error analysis.

To dispel these concerns, we quantitatively inspected the results for each questionable histogram. For example, consider the second histogram from top in the right column of Fig. 5. On this network, $|T - B|$ and $|B - T|$ were measured to be 694 and 476, respectively. Let us assume an inordinate worst case and suppose that every point in the histogram bins ranging all the way from $\mathcal{RE} = 2^0$ to 2^{20} was actually fixed and incorrectly classified as "not fixed" by our test. Additionally let us even suppose that each point in these bins was a distinct fixed point with no duplicates. Even then, these bins contain only 86 points, which cannot account for even half of the difference $|T - B| - |B - T|$. The same check was performed on every network with size $N \in \{24, 32, 48, 64\}$ (where it was claimed that $|T - B|$ was significantly larger than $|B - T|$), again using the generous cap of 2^{20}. On average, the number of points in the questionable bins was only 27.7% of $|T - B| - |B - T|$. So we can be quite confident that our results reported in [2] are essentially correct.

As justified empirically in the source paper [2], given two points $\overline{v}^{(1)}$ and $\overline{v}^{(2)}$ that had both been classified as fixed, they were considered duplicates only if

$$\max_i |\overline{v}_i^{(1)} - \overline{v}_i^{(2)}| < 2^{-21}. \tag{55}$$

Based on this test, we extract unique fixed points from a set with duplicates as follows. First, an adjacency graph is formed, where two points are adjacent if they were detected as duplicates. Next, the connected components of the graph are identified. Finally, one representative unique fixed point is chosen from each connected graph component.

References

[1] Gerald B Folland. *Advanced Calculus*. Prentice Hall, 2002.

[2] Garrett Katz and James Reggia. Towards global solution of the fixed point equations in recurrent neural networks. *Submitted*, 2017.

[3] Michael Taylor. The inverse function theorem via Newton's method. `http://www.unc.edu/math/Faculty/met/invfn.pdf`. Accessed: 2016-11-21.

[4] Wang Xinghua. Convergence of Newton's method and inverse function theorem in Banach space. *Mathematics of Computation of the American Mathematical Society*, 68(225):169–186, 1999.