Contractive non-autonomous systems

In the previous section, we found in the most general case that fixed points are unstable. For many neural net models, the goal is to train the network to converge toward an optimal point i.e. a fixed point. In order to guarantee this situation, we impose additional conditions on the training functions. For this reason, this section explores non-autonomous systems where all orbits converge to a fixed point or a periodic point.

To motivate the following theorem, suppose the training functions have the same fixed point of weights. The problem is to ensure that the iterates of the training process converge to this desired state.

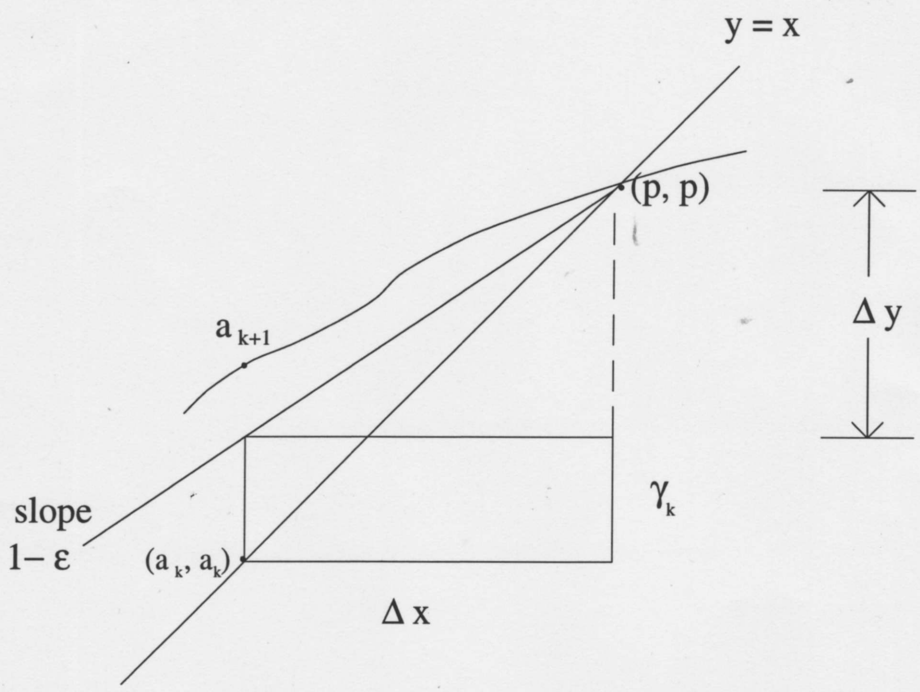
THEOREM 3.9. *Suppose $f_i : \mathbb{R} \longrightarrow \mathbb{R}$ is a sequence of differentiable functions. Further, assume that p is a fixed point of each f_i, i.e. $f_i(p) = p$ for every i. Suppose there exists a $\delta > 0$ and an $\epsilon > 0$ so that for any $x \in (p - \delta, p]$ we have that $0 \leq f_i'(x) < (1 - \epsilon)$ for all i. Then for any $x_0 \in (p - \delta, p]$, the orbit of x_0 converges to p, i.e. $\lim_{k \to \infty} f_k \circ f_{k-1} \circ \ldots f_2 \circ f_1(x_0) = p$*

Proof: Set $a_0 = x_0$. Set $a_1 = f_1(x_0)$. Set $a_2 = f_2 \circ f_1(x_0)$ $, \ldots, a_k = f_k \circ \cdots \circ$

$f_1(x_0)$.

We first need to show that $f_k \circ f_{k-1} \circ \ldots \circ f_2 \circ f_1(x_0)$ lies in $(p-\delta, p]$ for all k. We do this by induction. The base case $a_0 = x_0$ is in $(p-\delta, p]$ because that is the assumption in the theorem. For the inductive step, suppose $f_{k-1} \circ \ldots \circ f_2 \circ f_1(x_0)$ lies in $(p-\delta, p]$. The derivative $f_k{}'(x) \geq 0$ on $(p-\delta, p]$, which implies $a_k = f_k(a_{k-1}) \geq a_{k-1} > p - \delta$. Suppose $a_k > p$. The Mean Value Theorem implies there a point b in $[a_k, p]$ with $f_k{}'(b) < 0$. This contradicts the hypothesis, so $a_k \leq p$.

The next part involves showing that $a_k \to p$.



The diagram shows the fixed point line $y = x$, and the line with slope $1 - \epsilon$ that passes thru the point (p, p). Also, shown is the point (a_k, a_k), where a_k is the result of k iterations from the initial point x_0. The curve represents the graph of the function f_{k+1}. This curve must lie above the line with slope $1 - \epsilon$ because if not then the Mean Value Theorem tells us that there must be a point $a \in (p - \delta]$ where ${f_{k+1}}'(a) < 1 - \epsilon$.

This fact allows us to find a positive lower bound for $a_{k+1} - a_k$. To do this, we define a few variables using the geometry of the picture. Consider the vertical line segment that starts at the point (a_k, a_k) and ends where the vertical line intersects the line through (p, p) with slope $1 - \epsilon$. Call the length of this line segment, γ_k. Set $\Delta x = p - a_k$, and set $\Delta y = p - a_k - \gamma_k$, as shown in the picture. We see that $\frac{\Delta y}{\Delta x} = 1 - \epsilon$.

Substitute the expressions for Δy and Δx into the equation, $\frac{\Delta y}{\Delta x} = 1 - \epsilon$. Hence, $\gamma_k = (p - a_k) - (p - a_k)(1 - \epsilon) = (p - a_k)\epsilon$. Since the graph of the function f_{k+1} lies above the line with slope $1 - \epsilon$, we see that $a_{k+1} - a_k \geq \gamma_k = (p - a_k)\epsilon$.

This proves that $\lim_{k \to \infty} a_k = p$. To see why, let $\gamma > 0$. If $|p - a_k| \geq \gamma$ for all natural numbers k, then $a_{k+1} - a_k \geq \gamma\epsilon$ for each k. This is a contradiction since $\sup\{a_1, a_2, \dots\} \leq p$ and both ϵ and γ are positive. Hence there is an m so that $|p - a_m| < \gamma$. Since $\{a_k\}$ is increasing, then $|p - a_k| < \gamma$ for any $k \geq m$. ■

Notice that $a_{k+1} - a_k \geq (p - a_k)\epsilon$ allows us to estimate the rate of convergence of x_0 to p. The following is similar to the previous theorem, but not quite a generalization because there are no derivative hypotheses for metric spaces. The theorem is relevant because it helps determine when a training algorithm will converge to a unique training function.

THEOREM 3.10. *Suppose (X, d) is a complete metric space. Suppose $f_i : X \longrightarrow X$ is a sequence of contraction mappings such that there exists λ_i with $0 \leq \lambda_i < 1$ so* *that* $d(f_i(x), f_i(y)) \leq \lambda_i d(x,y)$ *for all* i *and for all* $x, y \in X$. *Suppose that* $\prod_{i=1}^{\infty} \lambda_i = 0$. *By the Contraction Mapping Lemma [SPIVAK], each* f_i *has a unique fixed point* p_i. *Suppose that all the fixed points are the same point i.e. there exists* p *so that* $p = p_i$ *for every* i. *Then* $\lim_{k\to\infty} f_k \circ f_{k-1} \ldots f_2 \circ f_1(x) = p$ *for any* $x \in X$.

PROOF: Let $x \in X$. The hypotheses imply that $d(f_1(x), p) = d(f_1(x), f_1(p)) \leq \lambda_1 d(x,p)$. Again, apply the hypotheses in the same way,
$d(f_2 \circ f_1(x), p) = d(f_2 \circ f_1(x), f_2(p)) \leq \lambda_2 d(f_1(x), p) \leq \lambda_2 \lambda_1 d(x,p)$.
Suppose by induction that $d(f_{k-1} \circ f_{k-2} \ldots f_1(x), p) \leq (\prod_{i=1}^{k-1} \lambda_i) d(x,p)$.
Set $y = f_{k-1} \circ f_{k-2} \ldots f_1(x)$.
Then $d(f_k \circ f_{k-1} \ldots f_1(x), p) = d(f_k(y), f_k(p)) \leq \lambda_k d(y,p) \leq (\prod_{i=1}^{k} \lambda_i) d(x,p)$, by the induction hypothesis. The hypothesis $\prod_{i=1}^{\infty} \lambda_i = 0$ proves that $\lim_{k\to\infty} f_k \circ f_{k-1} \ldots f_2 \circ f_1(x) = p$. ■

EXAMPLE:

We discuss two different examples. In the first one $\prod_{i=1}^{\infty} \lambda_i \neq 0$; in the second one $\prod_{i=1}^{\infty} \lambda_i = 0$. Consider the sequence of functions $f_n : \mathbb{R} \longrightarrow \mathbb{R}$ where $f_n(x) = s_n x + t_n$. By example 1.1, we have $\lim_{k\to\infty} f_k \circ f_{k-1} \ldots f_1(x) = (\prod_{i=1}^{\infty} s_i)x + \sum_{k=2}^{\infty} (\prod_{i=k}^{\infty} s_i) t_{k-1}$.

In our first example, we set $s_n = (1 - 2^{-n})x$ and set $t_n = 2^{-n}$. As in example 1.1, we obtain the following. When $x < 1$, we have $\lim_{n\to\infty} f_n \circ f_{n-1} \ldots f_1(x) < 1$, and when $x > 1$. we have $\lim_{n\to\infty} f_n \circ f_{n-1} \ldots f_1(x) > 1$. ■

In the second example, set $g_n(x) = s_n x + t_n$ where $s_n = 1 - \frac{1}{n+1}$ and $t_n = \frac{1}{n+1}$. As we calculated for $\{f_1, f_2, \ldots, \}$, we see that $\lim_{n\to\infty} g_n \circ g_{n-1} \ldots g_1(x) = (\prod_{i=1}^{\infty} s_i)x +$

$\sum_{k=2}^{\infty}(\prod_{i=k}^{\infty} s_i)t_{k-1} = \sum_{k=2}^{\infty}(\prod_{i=k}^{\infty} s_i)t_{k-1}$ because by Theorem 1.1, $(\prod_{i=1}^{\infty} s_i) = 0$. Since $g_n(1) = 1$ for all n, the orbit of the point 1 converges to 1 i.e. $\lim_{n\to\infty} g_n \circ g_{n-1} \dots g_1(1) = 1$. This implies that $\sum_{k=2}^{\infty}(\prod_{i=k}^{\infty} s_i)t_{k-1} = 1$. Thus, $\lim_{n\to\infty} g_n \circ g_{n-1} \dots g_1(x) = 1$ for any real number x. ■

Although these two examples appear to be similar, the difference is that the rate at which the slopes of the line approach 1, i.e. non-hyperbolicity, is different. This is why $\prod_{i=1}^{\infty} \lambda_i \neq 0$ for $\{f_1, f_2, \dots\}$, while $\prod_{i=1}^{\infty} \lambda_i = 0$ for $\{g_1, g_2, \dots\}$.

The next theorem is similar to the previous theorem; however, in Theorem 3.11, the fixed points p_i of each f_i no longer have to be the same. We only require that these fixed points p_i converge to a single point p. The theorem is relevant to training because it tells us that the training algorithm will converge to a unique point, p, in the weight space, as long as the fixed points, p_i, of the training functions f_i in the limit converge to p.

THEOREM 3.11. *Suppose* (X, d) *is a complete metric space. Suppose* $f_i : X \longrightarrow X$ *is a sequence of contraction mappings such that there exists* λ_i *with* $0 \leq \lambda_i < 1$ *so that* $d(f_i(x), f_i(y)) \leq \lambda_i d(x, y)$ *for all* i *and for all* $x, y \in X$. *By the Contraction Mapping Lemma, each* f_i *has a unique fixed point* p_i. *Suppose that* $\lim_{i\to\infty} p_i = p$ *and that* $\sup\{\lambda_i : i \geq 1\} = \lambda < 1$. *Then* $\lim_{k\to\infty} f_k \circ f_{k-1} \dots f_2 \circ f_1(x) = p$ *for any* $x \in X$.

Proof: Fix $x \in X$. Suppose $\sup\{\lambda_i : i \geq 1\} = \lambda < 1$. Let $\epsilon > 0$.

Set $\gamma = \min\{\frac{\epsilon(1-\lambda)}{3}\} \leq \frac{\epsilon}{3}$. By hypothesis, there exists N so that $d(p_n, p_m) < \gamma$ and $d(p_n, p) < \gamma$ whenever $n, m \geq N$. Set $y = f_N \circ f_{N-1} \circ \dots f_2 \circ f_1(x)$. Set

$L = d(p_{N+1}, y)$. The case $L = 0$ is a simple modification of the steps for the proof of $L > 0$. If $L > 0$, there exists Q so that $\lambda^k < \frac{\epsilon}{3L}$ whenever $k \geq Q$. Choose $M = N + Q$. When $n \geq M$ we have $d(f_n \circ f_{n-1} \circ \ldots f_1(x), p) = d(f_{N+j} \circ \cdots \circ f_{N+1}(y), p)$, for some $j \geq Q$. Hence,

$$
\begin{aligned}
& d(f_{N+j} \circ \cdots \circ f_{N+1}(y), p) \leq d(f_{N+j} \circ \cdots \circ f_{N+1}(y), p_{N+j}) + d(p_{N+j}, p) \\
& < (\prod_{i=N+1}^{N+j} \lambda_i) d(y, p_{N+1}) + (\prod_{i=N+1}^{N+2} \lambda_i) d(p_{N+1}, p_{N+2}) + \ldots \\
& + \lambda_{N+j-1}\lambda_{N+j} d(p_{N+j-2}, p_{N+j-1}) + \lambda_{N+j} d(p_{N+j-1}, p_{N+j}) + \tfrac{\epsilon}{3} \\
& \leq (\prod_{i=N+1}^{N+j} \lambda_i) d(y, p_{N+1}) + \gamma[\lambda_{N+j} + \lambda_{N+j}\lambda_{N+j-1} + \cdots + (\prod_{i=N+2}^{N+j} \lambda_i)] + \tfrac{\epsilon}{3} \\
& \leq \lambda^j L + \gamma[\tfrac{1}{1-\lambda}] + \tfrac{\epsilon}{3} < \epsilon. \blacksquare
\end{aligned}
$$

In the next theorem, we assume the non-autonomous system, $(X, \{f_1, f_2, \ldots f_n, f_1, f_2, \ldots f_n, \ldots\})$ has period n. This theorem is different from the previous theorems because if we know that our non-autonomous system is periodic, then we do not have to make any assumptions about where the fixed points of each of the training functions f_i are.

THEOREM 3.12. *Let (X, d) be a complete metric space. Suppose $f_i : X \longrightarrow X$ where $1 \leq i \leq n$ are continuous functions. Suppose there exists $\lambda_1, \ldots, \lambda_n \geq 0$ with $\prod_{i=1}^{n} \lambda_i < 1$, so that for any $x, y \in X$, we have $d(f_i(x), f_i(y)) \leq \lambda_i d(x, y)$ where $1 \leq i \leq n$.*

There exists $q \in X$ so that for any $x \in X$ $\lim_{k \to \infty} [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x) = q$. Further, the sequence of functions $\{g_1, g_2, \ldots\}$ approaches the following orbit of period n,

$\{q, f_1(q), f_2 \circ f_1(q), f_3 \circ f_2 \circ f_1(q), \ldots, f_{n-1} \circ f_{n-2} \circ \ldots f_2 \circ f_1(q)\}$. More precisely, for any $\epsilon > 0$ and for any $x \in X$ there exists N so that $m \geq N$ where $m = kn + r$ and $0 < r < n$ implies that $d(f_r \circ f_{r-1} \circ \ldots f_2 \circ f_1(x) \circ [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x)$, $f_r \circ f_{r-1} \ldots f_2 \circ f_1(q)) < \epsilon$. Notice that the case of $r = 0$ is already covered above.

Proof: Set $h = f_n \circ f_{n-1} \ldots f_2 \circ f_1$. First, we show that h is a contraction mapping. This assertion follows from the following inequality. The inequality is $d(h(x), h(y)) \leq \lambda_n d(f_{n-1} \circ \ldots f_2 \circ f_1(x), d(f_{n-1} \circ \ldots f_2 \circ f_1(y) \leq \lambda_n \lambda_{n-1} \ldots \lambda_2 \lambda_1 d(x, y)$. Now, apply the Contraction Mapping Lemma to deduce that h has a unique fixed point q. The Contraction Mapping Lemma implies that for any $x \in X$

$$\lim_{k \to \infty} [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x) = q.$$

This equation and the continuity of f_1 implies that

$$\lim_{k \to \infty} f_1 \circ [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x) = f_1(q).$$

Use the same argument on f_2:

$$\lim_{k \to \infty} f_2 \circ f_1 \circ [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x) = f_2 \circ f_1(q).$$

By induction we see that

$$\lim_{k \to \infty} f_r \ldots f_2 \circ f_1 \circ [f_n \circ f_{n-1} \circ \cdots \circ f_1]^k(x) = f_r \ldots f_2 \circ f_1(q),$$

where $0 < r < n$. ■