Erratum note for the techreport, *The "echo state" approach to analysing and training recurrent neural networks*

Herbert Jaeger
Jacobs University Bremen

January 24, 2010

Abstract

In the technical report *The "echo state" approach to analysing and training recurrent neural networks* from 2001, a number of equivalent conditions for the *echo state property* were given. As pointed out by Tobias Strauss, one of them is too weak and not equivalent to the others. Here I rectify this error, stating the correct version of that condition, which was suggested by Tobias Strauss.

1 Introduction

This erratum note is not a stand-alone article. It just provides a corrected version of Definition 3, Proposition 1, and the associated proofs from the technical report [1], without providing further explanation of context. Detecting the error, correcting it, and providing a new proof for the ensuing revised proposition, is all due to Tobias Strauss [2].

2 Corrected version of Definition 3

The version of Definition 3 in the original techreport provided three properties that were claimed in Proposition 1 to be all equivalent with the echo state property. However, the first property was too weak. Here a corrected version of this definition is given. The only change is in the statement of property 1. It was called the *state contracting* property in the original techreport. Tobias Strauss calls the corrected version the *uniformly state contracting* property, a terminology that

I would want to adopt (and dismiss the old name altogether with its defunct definition).

Definition 4 [Repeated from [1], with a correction in the statement of the *state contracting* property, which becomes the *uniformly state contracting* property.] *Assume standard compactness conditions and a network without output feedback.*

1. [Corrected] *The network is called* uniformly state contracting *if there exists a null sequence* $(\delta_h)_{h\geq 0}$ *such that for all right-infinite input sequences* $\bar{\mathbf{u}}^{+\infty}$*, and for all states* $\mathbf{x}, \mathbf{x}' \in A$*, for all* $h \geq 0$*, for all input sequence prefixes* $\bar{\mathbf{u}}_h = \mathbf{u}(n), \ldots, \mathbf{u}(n+h)$ *it holds that* $d(T(\mathbf{x}, \bar{\mathbf{u}}_h), T(\mathbf{x}', \bar{\mathbf{u}}_h)) < \delta_h$*, where* d *is the Euclidean distance on* $\mathbb{R}^N$*.*
2. *The network is called* state forgetting *if for all left-infinite input sequences* $\ldots, \mathbf{u}(n-1), \mathbf{u}(n) \in U^{-\mathbb{N}}$ *there exists a null sequence* $(\delta_h)_{h\geq 0}$ *such that for all states* $\mathbf{x}, \mathbf{x}' \in A$*, for all* $h \geq 0$*, for all input sequence suffixes* $\bar{\mathbf{u}}_h = \mathbf{u}(n-h), \ldots, \mathbf{u}(n)$ *it holds that* $d(T(\mathbf{x}, \bar{\mathbf{u}}_h), T(\mathbf{x}', \bar{\mathbf{u}}_h)) < \delta_h$*.*
3. *The network is called* input forgetting *if for all left-infinite input sequences* $\bar{\mathbf{u}}^{-\infty}$ *there exists a null sequence* $(\delta_h)_{h\geq 0}$ *such that for all* $h \geq 0$*, for all input sequence suffixes* $\bar{\mathbf{u}}_h = \mathbf{u}(n-h), \ldots, \mathbf{u}(n)$*, for all left-infinite input sequences of the form* $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}_h, \bar{\mathbf{v}}^{-\infty}\bar{\mathbf{u}}_h$*, for all states* $\mathbf{x}$ *end-compatible with* $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}_h$ *and states* $\mathbf{x}'$ *end-compatible with* $\bar{\mathbf{v}}^{-\infty}\bar{\mathbf{u}}_h$ *it holds that* $d(\mathbf{x}, \mathbf{x}') < \delta_h$*.*

The following identically re-states Proposition 1 from the techreport, except that *state contracting* has been changed to *uniformly state contracting.*

Proposition 1 *Assume standard compactness conditions and a network without output feedback. Assume that* T *is continuous in state and input. Then the properties of being uniformly state contracting, state forgetting, and input forgetting are all equivalent to the network having echo states.*

The following proof of Prop. 1 by and large replicates the original proof from the techreport, except a re-arrangement, some completions and adding a proof for the implication *uniformly state contracting* $\Rightarrow$ *echo states.*

Proof.

Part 1: *echo states* $\Rightarrow$ *uniformly state contracting.*

Let

$$D = \{(\mathbf{x},\mathbf{x}') \in A^2 \,|\, \exists\, \bar{\mathbf{u}}^\infty \in U^{\mathbb{Z}}, \exists\, \bar{\mathbf{x}}^\infty, \bar{\mathbf{x}}'^\infty \in A^{\mathbb{Z}}, \exists n \in \mathbb{Z} : \bar{\mathbf{x}}^\infty, \bar{\mathbf{x}}'^\infty \text{ compatible with } \bar{\mathbf{u}}^\infty \text{ and } \mathbf{x} = \bar{\mathbf{x}}(n) \text{ and } \mathbf{x}' = \bar{\mathbf{x}}'(n)\}$$

denote the set of all state pairs that are compatible with some input sequence. It is easy to see that the echo state property is equivalent to the condition that D contain only identical pairs of the form $(\mathbf{x},\mathbf{x})$.

Like in the original techreport, we first derive an alternative characterization of D. Consider the set

$$P^+ = \{(\mathbf{x},\mathbf{x}',1/h) \in A \times A \times [0,1] \;|\; h \in \mathbb{N}, \exists \bar{\mathbf{u}}^h \in U^h, \mathbf{x} \text{ and } \mathbf{x}' \text{ are end-compatible with } \bar{\mathbf{u}}^h\}.$$

Let D^+ be the set of all points $(\mathbf{x},\mathbf{x}')$ such that $(\mathbf{x},\mathbf{x}',0)$ is an accumulation point of P^+ in the product topology of $A \times A \times [0,1]$. Note that this topology is compact and has a countable basis. We show that $D^+ = D$.

$D \subseteq D^+$: If $(\mathbf{x},\mathbf{x}') \in D$, then $\forall h :\ (\mathbf{x},\mathbf{x}',1/h) \in P^+$ due to input shift invariance, hence $(\mathbf{x},\mathbf{x}',0)$ is an accumulation point of P^+ .

$D^+ \subseteq D$: (a) From continuity of T and compactness of A, a straightforward argument shows that D^+ is closed under network update T, i.e., if $(\mathbf{x},\mathbf{x}') \in D^+$, $\mathbf{u} \in U$, then $(T(\mathbf{x},\mathbf{u}), T(\mathbf{x}',\mathbf{u})) \in D^+$. (b) Furthermore, it holds that for every $(\mathbf{x},\mathbf{x}') \in D^+$, there exist $\mathbf{u} \in U$, $(\mathbf{z},\mathbf{z}') \in D^+$ such that $(T(\mathbf{z},\mathbf{u}), T(\mathbf{z}',\mathbf{u})) = (\mathbf{x},\mathbf{x}')$. To see this, let $\lim_{i\to\infty}(\mathbf{x}_i,\mathbf{x}'_i,1/h_i) = (\mathbf{x},\mathbf{x}',0)$. For each of the $(\mathbf{x}_i,\mathbf{x}'_i)$ there exist $\mathbf{u}_i$, $(\mathbf{z}_i,\mathbf{z}'_i) \in A \times A$ such that $(T(\mathbf{z}_i,\mathbf{u}_i), T(\mathbf{z}'_i,\mathbf{u}_i)) = (\mathbf{x}_i,\mathbf{x}'_i)$. Select from the sequence $(\mathbf{z}_i,\mathbf{z}'_i,\mathbf{u}_i)$ a convergent subsequence $(\mathbf{z}_j,\mathbf{z}'_j,\mathbf{u}_j)$ (such a convergent subsequence must exist because $A \times A \times U$ is compact and has a countable topological base). Let $(\mathbf{z},\mathbf{z}',\mathbf{u})$ be the limit of this subsequence. It holds that $(\mathbf{z},\mathbf{z}') \in D^+$ (compactness argument) and that $(T(\mathbf{z},\mathbf{u}), T(\mathbf{z}',\mathbf{u})) = (\mathbf{x},\mathbf{x}')$ (continuity argument about T). (c) Use (a) and (b) to conclude that for every $(\mathbf{x},\mathbf{x}') \in D^+$ there exists an input sequence $\bar{\mathbf{u}}^\infty$, state sequences $\bar{\mathbf{x}}(n)^\infty$, $\bar{\mathbf{x}}'(n)^\infty$ compatible with $\bar{\mathbf{u}}^\infty$, and $n \in \mathbb{Z}$ such that $\mathbf{x} = \mathbf{x}(n)$ and $\mathbf{x}' = \mathbf{x}'(n)$.

With this preparation we proceed to the proof of *echo states* $\Rightarrow$ *uniformly state contracting*, repeating (and translating to English) the argument given by Tobias Strauss.

Assume the network is not uniformly state contracting. This implies that for every null sequence $(\delta_i)_{i\geq 0}$ there exists a $h \geq 0$, an input sequence $\bar{\mathbf{u}}_h$ of length h, and states $\mathbf{x},\mathbf{x}' \in A$, such that

$$d(T(\mathbf{x},\bar{\mathbf{u}}_h), T(\mathbf{x}',\bar{\mathbf{u}}_h)) \geq \delta_h.$$

Since A is compact, it is bounded. Therefore, the sequence $(\mu_i)_{i\geq 0}$ defined by

$$\mu_i := \sup\{d(T(\mathbf{x}, \bar{\mathbf{u}}_i), T(\mathbf{x}', \bar{\mathbf{u}}_i)) \mid \mathbf{x}, \mathbf{x}' \in A, \bar{\mathbf{u}}_i \in U^i\}$$

is bounded, say by M. Because we assumed that the network is not uniformly state contracting, $(\mu_i)_{i\geq 0}$ is not a null sequence. Therefore there exists a subsequence $(\mu_{i_j})_{j\geq 0}$ of $(\mu_i)_{i\geq 0}$, which converges to some $\varepsilon > 0$. Since for all i, the space $U^i \times A$ is compact and $T : U^i \times A \to A$ is continuous, the supremum μ_i is realized by suitable $\mathbf{x}, \mathbf{x}' \in A$. Let $(\mathbf{x}_{i_j}, \mathbf{x}'_{i_j}) \in A^2$ be such that

$$(\mathbf{x}_{i_j}, \mathbf{x}'_{i_j}) \in \{(T(\mathbf{x}, \bar{\mathbf{u}}_{i_j}), T(\mathbf{x}', \bar{\mathbf{u}}_{i_j})) \mid \\ \bar{\mathbf{u}}_{i_j} \in U^{i_j}, \mathbf{x}, \mathbf{x}' \in A, d(T(\mathbf{x}, \bar{\mathbf{u}}_{i_j}), T(\mathbf{x}', \bar{\mathbf{u}}_{i_j})) = \mu_{i_j}\}.$$

Since A^2 is compact, there exist a subsequence $(\mathbf{x}_{i_{j_k}}, \mathbf{x}'_{i_{j_k}})_{k\geq 0}$ of $(\mathbf{x}_{i_j}, \mathbf{x}'_{i_j})_{j\geq 0}$ which converges to some $(\mathbf{y}, \mathbf{y}') \in A^2$. Obviously it holds that $(\mathbf{x}_{i_j}, \mathbf{x}'_{i_j}, \frac{1}{i_j}) \in P^+$. Thus $(\mathbf{y}, \mathbf{y}', 0)$ is an accumulation point of P^+, i.e., $(\mathbf{y}, \mathbf{y}') \in D^+$. On the other hand,

$$0 < \varepsilon = \lim_{k\to\infty} \mu_{i_{j_k}} = \lim_{k\to\infty} d(\mathbf{x}_{i_j} - \mathbf{x}'_{i_j}) = d(\mathbf{y}, \mathbf{y}').$$

This contradicts the echo state property, because D^+ does not contain pairs $(\mathbf{y}, \mathbf{y}')$ with $\mathbf{y} \neq \mathbf{y}'$.

Part 2: *uniformly state contracting $\Rightarrow$ state forgetting.*

Assume the network is not state forgetting. This implies that there exists a left-infinite input sequence $\bar{\mathbf{u}}^{-\infty}$, a strictly growing index sequence $(h_i)_{i\geq 0}$, states $\mathbf{x}_i, \mathbf{x}'_i$, and some $\varepsilon > 0$, such that

$$\forall i : \; d(T(\mathbf{x}_i, \bar{\mathbf{u}}^{-\infty}[h_i]), T(\mathbf{x}'_i, \bar{\mathbf{u}}^{-\infty}[h_i])) > \varepsilon,$$

where $\bar{\mathbf{u}}^{-\infty}[h_i]$ denotes the suffix of lenght h_i of $\bar{\mathbf{u}}^{-\infty}$. Complete every $\bar{\mathbf{u}}^{-\infty}[h_i]$ on the right with an arbitrary right-infinite input sequence, to get a series of right-infinite input sequences $(\bar{\mathbf{v}}_i)_{i=1,2,\ldots}$. For the i-th series $\bar{\mathbf{v}}_i$ it holds that $d(T(\mathbf{x}_i, \bar{\mathbf{v}}_i[h_i], T(\mathbf{x}'_i, \bar{\mathbf{v}}_i[h_i])) > \varepsilon$, where $\bar{\mathbf{v}}_i[h_i]$ is the prefix of length h_i of $\bar{\mathbf{v}}_i$, which contradicts the uniform state contraction property.

Part 3: *state forgetting $\Rightarrow$ input forgetting.*

Let $\bar{\mathbf{u}}^{-\infty}$ be a left-infinite input sequence, and $(\delta_h)_{h\geq 0}$ be an associated null sequence according to the state forgetting property. For the suffix $\bar{\mathbf{u}}_h$ of length h of $\bar{\mathbf{u}}^{-\infty}$, consider any pair $\mathbf{y}, \mathbf{y}'$ of states from A. By the state forgetting property it holds that $d(T(\mathbf{y}, \bar{\mathbf{u}}_h), T(\mathbf{y}', \bar{\mathbf{u}}_h)) < \delta_h$. Now consider any left-infinite $\bar{\mathbf{w}}^{-\infty}$ and $\bar{\mathbf{v}}^{-\infty}$. If, specifically, $\mathbf{y}, \mathbf{y}'$ are end-compatible with $\bar{\mathbf{w}}^{-\infty}$ and $\bar{\mathbf{v}}^{-\infty}$, respectively, it still holds that $d(T(\mathbf{y}, \bar{\mathbf{u}}_h), T(\mathbf{y}', \bar{\mathbf{u}}_h)) < \delta_h$. This implies that for all states $\mathbf{x}$ and $\mathbf{x}'$ which are end-compatible with $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}_h$ and $\bar{\mathbf{v}}^{-\infty}\bar{\mathbf{u}}_h$, respectively, it holds that $d(\mathbf{x}, \mathbf{x}') < \delta_h$.

Part 4: *input forgetting* $\Rightarrow$ *echo states.*

Assume that the network does not have the echo state property. Then there exists a left-infinite input sequence $\bar{\mathbf{u}}^{-\infty}$, states $\mathbf{x}, \mathbf{x}'$ end-compatible with $\bar{\mathbf{u}}^{-\infty}$, $d(\mathbf{x}$, such that $\mathbf{x}') > 0$. This leads immediately to a contradiction to input forgetting, by setting $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}_h = \bar{\mathbf{v}}^{-\infty}\bar{\mathbf{u}}_h = \bar{\mathbf{u}}^{-\infty}$.

References

[1] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science, 2001. http://www.faculty.jacobs-university.de/hjaeger/pubs/EchoStatesTechRep.pdf.

[2] T. Strauss. *Alternative Konvergenzmaße für die Beschreibung des Verhaltens von Echo-State-Netzen.* Diplomarbeit, Math.-Naturwissenschaftliche Fakultät, Institut für Mathematik, Universität Rostock, 2009.