Learning Oscillations Using Adaptive Control

Martin Georg Weiß
Graduiertenkolleg Technomathematik
Universität Kaiserslautern
Erwin-Schrödinger-Straße
67663 Kaiserslautern, Germany

September 8, 1997

Abstract

We study a model for learning periodic signals in recurrent neural networks proposed by Doya and Yoshizawa [7] that can be considered as a model for temporal pattern memory in animal motoric systems. A network receives an external oscillatory input and adjusts its weights so that this signal can be reproduced approximately as the network output after some time. We use tools from adaptive control theory to derive an algorithm for weight matrices with special structure. If the input is generated by a network of the same structure the algorithm converges globally and does not exhibit the deficiencies of the back-propagation based approach of Doya and Yoshizawa under a persistency of excitation condition. This simple algorithm can also be used for open loop identification under quite restrictive assumptions.

The persistency of excitation condition cannot be proven even for the matrices with special structure but for a 3d system. For higher dimensional systems we give connections to the theory of linear time-varying systems where this condition is generically tue (under assumptions which are also needed in the time-invariant case). However we cannot show that the linearized system related to the nonlinear neural network fulfils these generic assumptions.

1 Introduction

In [7] the following model for learning of motions was proposed: Trajectories of motion are assumed to be stored in some parts of the motor nervous system. Whenever we try to memorize a new motion, e. g. riding a bicycle or swimming, we achieve our goal by conscious repetition in a way of supervised learning. During this process some high level components of the nervous systems adjust or influence lower level components such that the difference between the trajectory that should be generated and the one that is actually produced is minimized. This constitutes an on-line learning method. After a while we can repeat the desired trajectory seemingly without thinking: The lower level components can autonomously generate the signals necessary for the control of the muscles without supervision.

Based on these assumptions a neural network was suggested that operates in two modes: In the first mode the network receives a periodic signal to be learned. The weights are adapted by recurrent backpropagation with the usual squared difference between the reference signal and the network output as the minimization criterion. After some learning period, when the error is sufficiently small, the weight adaptation is cut off and the network operates in the second mode: The output replaces the reference signal as the input so that the network is running autonomously now. If learning was successful the network should be able to sustain an output oscillation that approximates the reference signal.

The results in [7] show that simple wave forms can be learned and sustained. Signals containing more than one period can not be sustained. However, no theoretical explanation could be given whether this was due to the well-known deficiencies of the recurrent back-propagation algorithm (like the use of approximate partial derivatives, see e. g. [13]) or to some intrinsic properties of the problem. Accordingly no proof of convergence could be given.

In this paper we propose a learning algorithm based on techniques from adaptive control theory that sheds more light on the problem and allows for easy convergence speedup. Nevertheless, the control law uses only structures that can be interpreted as neural networks. No linear filters are required. In Section 2 we introduce the model and mathematical problem formulation. Then our learning rule is formulated with the proof of convergence in Section 3 and implications for the learning problem are given. Numerical results are shown in Section 4. Whereas the tracking goal is achieved in any case the convergence of the weights requires a persistency of excitation condition whose practical relevance is studied in Section 5. We also hint at a similar learning law for discrete time neural networks in Section 6.

2 Problem Formulation

We consider additive Hopfield network with state $x \in \mathbb{R}^m$, scalar input u, scalar output y and adjustable weights $W(t) \in \mathbb{R}^{m \times m}$:

$$\begin{aligned} \dot{x} &= -\tau x + W(t)\sigma(x) + b\sigma(u) \\ y &= c^t x \end{aligned} \tag{1}$$

Only the output of the system can be observed. We make the

Assumption 2.1 *The sigmoidal activation function $\sigma \in C^1(\mathbb{R})$ fulfills:*

$$\sigma(-x) = -\sigma(x) \text{ for all } x \in \mathbb{R} \tag{2}$$

$$\|\sigma\|_\infty < \infty \tag{3}$$

$$\|\sigma'\|_\infty < \infty \tag{4}$$

$$\sigma'(x) > 0 \text{ for all } x \in \mathbb{R} \tag{5}$$

So σ is a real analytic bounded odd function with bounded derivative like the main examples in neural networks, $\sigma = \tanh$ and $\sigma = \arctan$. For some results only a subset of thes assumptions is required.

The system receives a periodic input signal u that it should learn in the following sense: The weights W are adjusted by a differential equation so that the output y asymptotically tracks the reference signal u.

In order to be able to replicate u exactly we assume that u is generated by a network of the same structure running in closed loop:

$$\begin{aligned} \dot{x}^\star &= -\tau x^\star + W^\star\sigma(x^\star) + b\sigma(y^\star) \\ y^\star &= c^t x^\star \end{aligned} \tag{6}$$

Motiveted by [14] we choose a sparse structure in dimension $n+1=m$ where not all weights can be adjusted.

$$W = A(w) = \begin{bmatrix} J_n(-\alpha) & 0_{n\times 1} \\ w^t & -\alpha \end{bmatrix}, \quad b = \mathbf{e}_n, \quad c = \mathbf{e}_{n+1} \tag{7}$$

Here $J_n(-\alpha)$ denotes a Jordan block of size n and $w \in \mathbb{R}^n$ corresponding to self-inhibitory connections for $\alpha > 0$. We assume that α is fixed and known; only $w_1, \ldots, w_n$ can be adjusted.

The slighly more general structure

$$A(w) = \begin{bmatrix} -\alpha_1 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & 1 & \\ & & & -\alpha_n & 0 \\ w_1 & \ldots & \ldots & w_n & -\alpha_{n+1} \end{bmatrix} \tag{8}$$

with fixed but known $\alpha_i \in \mathbb{R}$ can be handled as well but for simplicity we first assume (7) and give an extension in Theorem 3.5.

The choice of structure (7) is motivated by the following: The unit vector c enables us to write the closed loop system as an autonomous Hopfield system without inputs

$$\dot{x}^\star = -\tau x^\star + (A(w^\star) + bc^t)\sigma(x^\star) \ . \tag{9}$$

It is easily seen that we can choose n eigenvalues of $A(w^\star) + bc^t$ symmetric to the real axis by appropriate choice of $w^\star$ since the characteristic polynomial of $\tilde{A}(w) := A(w) + bc^t$ is

$$\chi_{\tilde{A}(w)}(\lambda) = (\lambda + \alpha)^{n+1} - (w_n(\lambda + \alpha)^{n-1} + \ldots + w_2(\lambda + \alpha) + w_1) \tag{10}$$

So we can choose n eigenvalues $\lambda_1, \ldots, \lambda_n$ arbitrarily symmetric to the real axis, the last eigenvalue results as

$$\lambda_{n+1} = \mathrm{trace}(\tilde{A}(w)) - \sum_{i=1}^{n} \lambda_i$$

Eigenvalue conditions are a main ingredient in bifurcation theory which we use to generate periodic orbits. For suitable values of $w^\star$ (and stronger continuity assumptions on σ) a Hopf bifurcation occurs around the fixed point 0 of (9) yielding an asymptotically stable periodic orbit, see [18]. As the proof for the learning rule do not require more smoothness we do not pursue this further and make the

Assumption 2.2 *The system (9) has a nontrivial periodic solution yielding an asymptotically stable orbit.*

A nontrivial periodic solution in such a system will always give non-constant output:

Lemma 2.3 *A periodic solution of the system (6) is constant iff one component is constant.*

Proof: The proof is based on the simple fact that a periodic solution of an one-dimensional autonomous ordinary differential equation $\dot{x} = g(x)$ is necessarily constant, i. e. trivially periodic. Let $x(t)$ be a periodic solution of (6) with $x_i(t) = c_i$ constant for some $i \in \{1, \ldots, n+1\}$. If $i > 1$ the equation for component $i - 1$ is autonomous:

$$\dot{x}_{i-1} = -\tau x_{i-1} - \alpha\sigma(x_{i-1}) + \sigma(c_i)$$

So the periodicity of x_{i-1} implies that x_{i-1} is constant as well. Inductively we conclude that $x_1, \ldots, x_i$ are constant. If $i < n + 1$ holds we argue using the equation of component i. Since x_i is constant we have for all t

$$0 = \dot{x}_i(t) = -\tau c_i - \alpha\sigma(c_i) + \sigma(x_{i+1}(t))$$

Because of bijectivity of σ by (5) component x_{i+1} is constant as well. Again, inductively we get the result. $\circ$

As a consequence the output $y(t) = x_{n+1}(t; x_0)$ of a nontrivial periodic solution cannot be constant. From now on "periodic solution" will always mean "nontrivial periodic solution". A similar structure has also been used in [14] where the linear case $\sigma = \text{id}$ was studied. On the other hand it is well known that - unlike the linear case - this structure reminiscent of normal forms in linear control theory cannot produce all dynamics which are possible for fully variable weight matrix W (see [2]).

In the proof of Theorem 3.3 it will become clear why we choose this special structure of the weight matrix instead of the simpler matrix

$$\tilde{A}(w) = \begin{bmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & 1 & \\ & & & \ddots & 0 \\ w_1 & \dots & \dots & w_n & w_{n+1} \end{bmatrix}$$

yielding a closed loop system with weight matrix in companion form.

Now our learning problem can be stated as follows: Find an adaptation rule f for w such that for all initial values $w(0) \in \mathbb{R}^n$, $x(0) \in \mathbb{R}^{n+1}$ the system

$$\begin{aligned} \dot{x}^\star &= -\tau x^\star + (A(w^\star) + bc^t)\sigma(x^\star) \\ \dot{x} &= -\tau x + A(w)\sigma(x) + b\sigma(y^\star) \\ \dot{w} &= f(w, y^\star, x) \end{aligned} \tag{11}$$

achieves tracking and parameter convergence, provided that $x^\star$ is periodic:

$$\lim_{t\to\infty} (y(t) - y^\star(t)) = 0 \tag{12}$$

$$\lim_{t\to\infty} (w(t) - w^\star) = 0 \tag{13}$$

3 Learning Rule

Consider the first mode of the model, the learning phase. Define the state error $e := x - x^\star$ and the parameter error $p := w - w^\star$. In the linear case $\sigma = \text{id}$ of [14] the proof for the learning rule $\dot{w} = \dot{p} = -(y - y^\star)Q\sigma(\underline{x})$ (where we abbreviate $\underline{x} = [x_1, \dots, x_n]^t$) is based on the fact that the function $V(e,p) = \frac{1}{2}(e^t Pe + p^t Q^{-1} p)$ is a Lyapunov function for (11) in the case $\tau + \alpha > 0$ for some positive definite P satisfying the Lyapunov equation $PA(w^\star) + A(w^\star)^t P = -S$ for arbitrary definite matrix $S \in \mathbb{R}^{n+1\times n+1}$. As a common phenomenon in adaptive control this is, however, not a strict Lyapunov function but only satisfies $\dot{V}(e,p) \leq -e^t Se$ This gives $e \to 0$ using standard arguments. The convergence proof for the weigths can be accomplished using the theory of *persistent excitation* or the method of *averaging*, see [14].

In analogy to [14] we propose the parameter adaptation law

$$\dot{w} = \dot{p} = -(y - y^\star)Q\sigma(\underline{x}) \tag{14}$$

where $Q \in \mathbb{R}^{n,n}$ is an arbitrary positive definite matrix. For $Q = \eta I$ the rule is local. Similarly to [14] this rule can be motivated by a time-varying quadratic Lyapunov function candidate

$$V(e,p) = \frac{1}{2}(e^t P(t)e + p^t Q^{-1} p) \tag{15}$$

Given solutions $x^\star$, x of (11) on $\mathbb{R}$ (with any learning rule giving solutions on $\mathbb{R}$) we derive an o.d.e. for the error that enables us to use linear theory:

$$\begin{aligned}
\dot{e} &= -\tau(x - x^\star) + A(w^\star)(\sigma(x) - \sigma(x^\star)) + (A(w) - A(w^\star))\sigma(x) \\
&= -\tau(x - x^\star) + A(w^\star)(\sigma(x) - \sigma(x^\star)) + \begin{bmatrix} 0 & 0 \\ p^t & 0 \end{bmatrix} \sigma(x) \\
&= (-\tau I + A(w^\star)\mathrm{diag}(\sigma'(\xi_i(t))))\ e + cp^t\sigma(\underline{x}) \\
&=: \tilde{A}(t)e + cp^t\sigma(\underline{x})
\end{aligned} \tag{16}$$

The functions ξ_i are determined by the mean value theorem

$$\sigma(x_i) - \sigma(x_i^\star) = \sigma'(\xi_i)(x_i - x_i^\star)$$

The error e is a special solution of (16); by this construction properties of (11) can be shown considering all o.d.e.'s (16) constructed from solutions $x^\star$ and x. We need the following two properties of $\tilde{A}$:

Lemma 3.1 *Consider $\sigma \in C^1(\mathbb{R})$ with Assumptions (3), (4) and (5). Let $\tau > -\alpha\|\sigma'\|_\infty$. Given C^1 functions $x, x^\star : \mathbb{R} \to \mathbb{R}^{n+1}$ and $w^\star \in \mathbb{R}^n$ the LTV $\dot{e} = \tilde{A}(t)e$ with $\tilde{A}(t) = -\tau I + A(w^\star)diag(s_k(t))$ where*

$$s_k(t) = \begin{cases} \frac{\sigma(x_k(t)) - \sigma(x_k^\star(t))}{x_k(t) - x_k^\star(t)}, & x_k(t) \neq x_k^\star(t) \\ \sigma'(x_k^\star(t)), & x_k(t) = x_k^\star(t) \end{cases}$$

is exponentially stable.

Proof: By the mean value theorem we have $\tilde{A}(t) = -\tau I + A(w^\star)\mathrm{diag}(\sigma'(\xi_k(t)))$ for continuous functions xi_k. So $\tilde{A}$ is continuous and unique solutions exist on $\mathbb{R}$. Denote $\kappa := \tau + \alpha\|\sigma'\|_\infty > 0$ and $\phi_k(t, t_0) = \exp \int_{t_0}^t -(\tau + \alpha s_k(\tilde{t}))dt$ the transition function of

$$\dot{e}_k = \tilde{a}_{kk}(t)e_k = -(\tau + \alpha s_k(t))e_k$$

satisfying $|\phi_k(t, t_0)e_k(0)| \le \exp(-\kappa(t - t_0))\,|e_k(t_0)|$. Since component n of the LTV $\dot{e} = \tilde{A}(t)e$ depends on e_n alone we get $|e_n(t)| \le C_n \exp(-\kappa t)\,|e_n(0)|$. with $C_n = 1$ and $\kappa_n = \kappa$. Component $n-1$ of the LTV reads $\dot{e}_{n-1} = -(\tau + \alpha s_{n-1}(t))e_{n-1} + s_n e_n$ giving

$$\begin{aligned}
|e_{n-1}(t)| &= \left| \phi_{n-1}(t,0)e_{n-1}(0) + \int_0^t \phi_{n-1}(t,\tilde{t})s_n(\tilde{t})e_n(\tilde{t})d\tilde{t} \right| \\
&\le \exp(-\kappa t)\,|e_{n-1}(0)| + \int_0^t \exp(-\kappa(t-\tilde{t}))\|\sigma'\|_\infty \exp(-\kappa\tilde{t})\,|e_n(0)|\,d\tilde{t} \\
&= \exp(-\kappa t)(|e_{n-1}(0)| + t\exp(-\kappa t)\|\sigma'\|_\infty\,|e_n(0)|) \\
&\le C_{n-1}\exp(-\kappa_{n-1}t)\,\|(e_{n-1}, e_n)(0)\|
\end{aligned}$$

for some $C_{n-1} > 0$,$0 \le \kappa_{n-1} \le \kappa_n$. Inductively we get by summing up $\|\underline{e}(t)\| \le C\exp(-\tilde{\kappa}t)\,\|\underline{e}(0)\|$. Since

$$\dot{e}_{n+1} = s_{n+1}e_{n+1} + \underline{e}^t w^\star$$

we now get exponential stability for the $n+1$-dimensional system as in the above since the inhomogneity $\underline{e}^t w^\star$ decays exponentially. ∘

As a generalization of Lemma 3.8 in [14] we have

Lemma 3.2 *Let $R : \mathbb{R} \to PD(n+1)$, $R \in L^\infty S$, be partitioned as $R = \begin{bmatrix} R_{11} & r_{12} \\ {r_{12}}^t & r_{22} \end{bmatrix}$. Define $\tilde{A} : \mathbb{R} \to \mathbb{R}^{n+1 \times n+1}$ as in Lemma 3.1. Then there exists a unique uniformly positive definite solution $P : \mathbb{R} \to PD(n+1)$ of*

$$\dot{P} + \tilde{A}P + P\tilde{A} = -R \qquad (17)$$

$$Pc = c \qquad (18)$$

iff $r_{12} = -\mathrm{diag}_{k=1}^{n}(s_k)w^\star$, $r_{22} = -2(\tau + \alpha s_{n+1})$ and ${r_{12}}^t R_{11}^{-1} r_{12} < r_{22}$ and R is uniformly positive definite. In that case P can be partitioned into

$$P = \begin{bmatrix} P_{11} & 0 \\ 0^t & 1 \end{bmatrix}$$

with $P_{11} : \mathbb{R} \to PD(n)$. Such matrix functions R always exist.

Proof: Let $\tilde{A}$ and P uniformly positive definite be given and suppose that (17) and (18) hold. By Krasovskii's theorem (17) holds (with uniqueness of the solution) iff $\tilde{A}(t)$ is exponentially stable. It remains to prove (18). By (7) $c = e_{n+1}$ so $Pc = c$ and symmetry of P imply the above partitioning. Partition

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{a}_{21}^t & \tilde{a}_{22} \end{bmatrix}$$

where $\tilde{A}_{11} = A(w^\star)\mathrm{diag}_{k=1}^{n}(s_k)$, $\tilde{a}_{21} = \mathrm{diag}_{k=1}^{n}(s_k)w^\star$ and $\tilde{a}_{22} = -\tau - \alpha s_{n+1}$. Now writing $\dot{P} + \tilde{A}P + P\tilde{A} = -R$ partitioned

$$\begin{bmatrix} \dot{P}_{11} + P_{11}\tilde{A}_{11} + \tilde{A}_{11}^t P_{11} & \tilde{a}_{21} \\ \tilde{a}_{21}^t & 2\tilde{a}_{22} \end{bmatrix} = \begin{bmatrix} R_{11} & r_{12} \\ {r_{12}}^t & r_{22} \end{bmatrix}$$

we have only the above choice for r_{12} and r_{22}. The condition ${r_{12}}^t R_{11}^{-1} r_{12} < r_{22}$ is well known to be equivalent to the positive definitenes of R which is uniform because of the assumption on P.

Conversely, if the above (in)equalities hold, one can easily verify (17) and (18).

It remains to show that a $R \in L^\infty$ exists. Since $r_{12}, r_{22} \in L^\infty$ because of $s_k \in L^\infty$ for all k. Choosing R_{11} large enough we get ${r_{12}}^t R_{11}^{-1} r_{12} < r_{22}$. Because of the boundedness properties of r_{12} and r_{22} we can choose $R_{11} \in L^\infty$. ∘

The function R will never be specified but we will only use the existence of suitable P. The equations in Lemma 3.2 are reminiscent of the Kalman-Yacubovich-Lemma which is closely related to positive realness and convergence of gradient-based algorithms (as our can also be seen) but we have not pursued this further. However, one can see by a coordinate transormation that $c = e_{n+1}$ can be replaced by any time-varying vector in the lemma. We give a convergence proof for (14) similar to the persistency of excitation proof in [14].

Theorem 3.3 *Consider (11) with $\sigma \in C^1$ satisfying (4). Let $x^\star$ be nontrivially periodic and $\tau > -\alpha\|\sigma'\|_\infty$. Let $x(0)$ and $w(0)$ be given. Then (14) solves the output tracking problem (12) and also the state error converges to 0. If $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ are linearly independent which is equivalent to the PE property of $\sigma(\underline{x^\star})$ then the state and parameter error converge to zero exponentially. Otherwise the weight vector converges to an affine subspace containing $w^\star$:*

$$w(t) \quad \to \quad w^\star + \{\sigma(\underline{x^\star}(t)) : t \in \mathbb{R}\}^\perp \qquad (19)$$

Proof: The right hand side of (11) with f from (14) is linearly bounded so solutions exist on $\mathbb{R}^+$ (see [3], Theorem 7.8). By Lemma 3.2 the derivative of (15) along (11) then is

$$\begin{aligned}\dot V(e,p) &= \frac{1}{2}e^t(\dot P + \tilde A P + P\tilde A^t)e + e^t P c p^t \sigma(\underline{x}) + p^t Q^{-1}\dot p \\ &= \frac{1}{2}e^t S e + e^t c p^t \sigma(\underline{x}) - (y-y^\star)p^t Q^{-1} Q \sigma(\underline{x}) \\ &= \frac{1}{2}e^t S e \le 0\end{aligned}$$

since $e^t c = (y - y^\star)$. The LaSalle invariance principle gives $e \to 0$. (or: As

$$0 \le \int_0^t e(\tilde t)^t S e(\tilde t) d\tilde t = V(0) - V(t) < \infty$$

we have $e \in L^2$. Since $x^\star$ is periodic and therefore bounded, x is bounded as a solution of the o.d.e.

$$\dot x = -\tau x + A(w)\sigma(x) + b\sigma(y^\star)$$

that is a stable LTI with bounded inhomogeneity and so e is bounded. By (16) also $\dot e$ is bounded. By Barbalat's Lemma we get $e \to 0$.)

We write a combined error system in the form

$$\begin{bmatrix}\dot e \\ \dot p\end{bmatrix} = \begin{bmatrix} & \tilde A & \begin{matrix}0_{n\times n} \\ \sigma(\underline{x^\star})^t\end{matrix} \\ 0_{n\times n} & -Q\sigma(\underline{x^\star}) & 0\end{bmatrix}\begin{bmatrix} e \\ p\end{bmatrix} \tag{20}$$

By Theorem 5.11 this system is exponentially stable iff $\tilde A$ is exponentially stable and $\sigma(\underline{x})$ is persistently exciting (PE for short). We know that the difference $\underline{x} - \underline{x^\star} = \underline{e}$ is an exponentially decaying function. This enables us to study $\underline{x^\star}$ instead of $\underline{x}$: As σ' is bounded by (4) the function σ itself is uniformly continuous. Thus

$$\sigma(\underline{x}) - \sigma(\underline{x^\star}) = diag\sigma'(\xi_i)(\underline{x} - \underline{x^\star})$$

is exponentially decaying as well and therefore an L^2 function on $\mathbb{R}^+$. By Lemma 5.8 it suffices to show that $\sigma(\underline{x^\star})$ is PE. As $\sigma(\underline{x^\star})$ is periodic Lemma 5.9 gives that the PE property is equivalent to the linear independence of $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$.

The partial convergence theorem (Theorem 2.7.4 in [15]) gives (19). ∘

Remark 3.4 *A short look at system (9) shows that the linear independence (LI for short) of the functions $\sigma(x_1), \ldots, \sigma(x_n)$ is a necessary condition for the uniqueness of the reference weights: For all $a \in \mathbb{R}^n$ with $a^t\sigma(\underline{x^\star}) = 0$ the networks with weights $w^\star$ and $w^\star + a$ can produce the same trajectories since the weights influence the last component only via*

$$\sum_{i=1}^n {w^\star}_i \sigma(x_i^\star) = {w^\star}^t\sigma(\underline{x^\star}) = (w^\star + a)^t \sigma(\underline{x^\star}) = \sum_{i=1}^n ({w^\star}_i + a_i)\sigma(x_i^\star)$$

On the other hand LI not necessarily guarantees weight convergence; some additional condition like (almost) periodicity is needed.

Looking at the proofs we see that tracking and even state error convergence follows from the stability of $\tilde A$. Repeating litterally the proofs of Lemma 3.2 and Theorem 3.3 we generalize

Theorem 3.5 *Consider the system (11) with* $A(w) = \begin{bmatrix} J & 0 \\ w^t & \alpha \end{bmatrix}$, $0 \neq b \in \mathbb{R}^{n+1}$ *arbitrary and* $\sigma \in C^1$ *with (3) and (4). Assume that for all measurable functions* $s_i : \mathbb{R} \to [\inf_{x\in\mathbb{R}} \sigma'(x), \sup_{x\in\mathbb{R}} \sigma'(x)]$ *the matrix function* $A(w^\star)diag(s_i)$ *is exponentially stable. Let* $x^\star$ *be nontrivially periodic and* $\tau > -\alpha\|\sigma'\|_\infty$. *Then the conclusions of Theorem 3.3 hold.*

Remark 3.6 *Constructions and proofs for periodic solutions in these more general structure are much more complicated. Even the proof of exponential stability of* $\tilde{A}(t)$ *usually is impossible since no closed form of the fundamental solution can be given. In some cases, like* J *upper triangular instead of* $J(-\alpha)$ *in (7) with* $J_{ii} < \tau\|\sigma'\|_\infty$, *sign conditions suffice. This property can be shown based on the structure of the matrix similar to structure (7) in the proof of Lemma 3.1.*

On the other hand we only need exponential stability for such functions s_i *that originate from periodic solutions of (11). There seems little sense in studying the difference.*

3.1 Open Loop Identification

So far the unknown weights have entered one row of the matrix only. If we allow fully unknown weight matrices $A(w) = A(W) = W$ for

$$\dot{x}^\star = -\tau x^\star + W^\star \sigma(x^\star) \tag{21}$$

then Theorem 3.3 fails to single out matrices W for which we achieve our aims: The Lyapunov function candidate for e and $p := W - W^\star$

$$V(e,p) = \frac{1}{2}(e^t P(t) e + \mathrm{trace}(pQ^{-1}p))$$

gives via

$$\begin{aligned} \dot{V}(e,p) &= \frac{1}{2}e^t(\dot{P} + \tilde{A}P + P\tilde{A}^t)e + e^t P p \sigma(x) + tracep^t Q^{-1}\dot{p} \\ &= \frac{1}{2}e^t S e + \mathrm{trace}(e^t P p \sigma(x)) + \mathrm{trace}(\dot{p}^t Q^{-1} p) \\ &= \frac{1}{2}e^t S e + \mathrm{trace}(\sigma(x) e^t P p) + \mathrm{trace}(\dot{p}^t Q^{-1} p) \end{aligned}$$

the learning rule

$$\dot{W} = -QPe\sigma(x) \tag{22}$$

This rule cannot be realized - even if the states are accessible - since P depends on $\tilde{A}$ and is therefore unknown. Besides, exponential stability of $-\tau I + W^\star \mathrm{diag}(s_k)$ is needed. The only practical estimate to guarantee this is

$$\|\sigma'\|_\infty \|W^\star\|_2 < \tau \tag{23}$$

However, this condition is obviously sufficient for the exponential stability of the LTV and therefore for the global asymptotic stability of the then unique fixed point 0 of system (21) (this is well known in the neural network community, see [4].

Nevertheless the above learning rule can be used for open loop identification using a controlled reference system and model.

Theorem 3.7 *Consider the system*

$$\begin{aligned} \dot{x}^\star &= -\tau x^\star + W^\star \sigma(x^\star) + b\sigma(u) \\ \dot{x} &= -\tau x + W\sigma(x) + b\sigma(u) \end{aligned} \tag{24}$$

with σ as in Theorem 3.3 and $W^\star$ satisfying (23). Then the leaning rule (22). solves the output tracking problem (12). If $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ are linearly independent then the state and parameter error converge to zero exponentially. Otherwise the weight matrix converges to an affine subspace containing $W^\star$:

$$W(t) \;\to\; W^\star + \left\{ W \in \mathbb{R}^{m\times m} : W_{i\cdot}{}^t \perp \{\sigma(\underline{x}^\star(t)) : t \in \mathbb{R}\} \text{ for all } i = 1, \ldots, m \right\}$$

Proof: The Lyapunov function candidate $V(e,p) = \frac{1}{2}(e^t P e + \mathrm{trace}(p^t Q^{-1} p))$ gives the above learning rule since the control terms cancel out. ∘

3.2 Convergence Speedup

The gain matrix Q can be made arbitrarily large (in the sense $Q > \eta I$, η large) but numerical simulations show that the convergence speed saturates for large η. The speed of convergence can be increased by using time-varying gains $R(t)^{-1}$ instead of Q in a learning rule

$$\dot{w} = \dot{p} = -(y - y^\star) R(t)^{-1} \sigma(\underline{x}) \tag{25}$$

where R is adapted on-line according to the differential equation

$$\dot{R} = \begin{cases} 0, & 0 \le t \le t_0 \\ \frac{1}{t}(-R(t) + \sigma(\underline{x})\sigma(\underline{x})^t), & t \ge t_0 \end{cases} \tag{26}$$

This can be motivated by an averaging analysis similar to [14]). Here $t_0 > 0$ is necessary to avoid the singularity in the right hand side. Denote by $\tilde{\phi}(t, t_0, z_0) = \phi(t, t_0) z_0 = \frac{t_0}{t} z_0)$ the solution of $\dot{z} = -\frac{1}{t} z$, $t \ge t_0 > 0$, $z(t_0) = z_0$. For arbitrary positive definite initial values $R(t_0)$ the solution

$$R(t) = \phi(t, t_0) R(t_0) + \int_{t_0}^{t} \phi(t, \tau) \sigma(\underline{x}(\tau)) \sigma(\underline{x}(\tau))^t d\tau$$

of (26) consist of an positive definite term due to initial conditions and a positive semidefinite term. As $\lim_{t\to\infty} \phi(t, t_0) = 0$ the solutions approach the autocovariance of $\sigma(\underline{x}^\star)$

$$R^\star := \lim_{t\to\infty} R(t) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \sigma(\underline{x}^\star(s)) \sigma(\underline{x}^\star(s))^t ds \tag{27}$$

which is positive definite iff $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ are linearly independent. Usually the gain matrix expresses some idea of the desired speed of adaptation which is modified by (26). This can be remedied by the law

$$\dot{w} = \dot{p} = -\eta(y - y^\star) \left\| R(t)^{-1} \right\|^{-1} R(t)^{-1} \sigma(\underline{x}) \tag{28}$$

with $\eta > 0$ that holds the norm of the gain matrix constant.

Theorem 3.8 *Under the assumptions of Theorem 3.3 and LI of $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ the modified learning laws (25), (26) and (28), (26) give exponential state and parameter convergence for all $R(t_0) \in PD(n)$ (with solutions considered on $[t_0, \infty)$).*

Proof: Solutions of the o.d.e.'s exist on $\mathbb{R}$ as R obeys a LTV with continuous r.h.s. (which does not depend on the weight dynamics!). Since $R(t_0) \in \mathrm{PD}(n)$ and $\lim_{t\to\infty} R(t) = R^\star \in \mathrm{PD}(n)$ by the assumptions continuity of $R(t)$ gives uniform positive definiteness and boundedness of R. For (28) we can therefore invert $R(t)$. So the equations (11) are still linearly bounded.

The proof of Theorem 3.3 still holds since Theorem 5.11 also holds for time-varying uniformly positive definite and bounded gains. ∘

3.3 Replication phase

We now consider the replication phase for system (11). After some time, t_0 say, the weight adaptation is switched off with weights now constant $w(t_0)$ and the learning system is fed its own output y:

$$\begin{aligned} \dot{x}^\star &= -\tau x^\star + (A(w^\star) + bc^t)\sigma(x^\star) \\ \dot{x} &= -\tau x + (A(w(t_0) + bc^t)\sigma(x) \end{aligned} \tag{29}$$

The learning system is running independent from the referrence system now. We investigate in what sense x approximates $x^\star$.

Consider generally (21) with an asymptotically stable orbit and denote $\lambda_0 := W^\star \in \mathbb{R}^{m^2}$ for correspondence with the cited literature. In the proof of Theorem 27.11 in [3] it is shown that the closed orbits $\gamma_0 := \gamma(\lambda_0) = \{x_{\lambda_0}(t) : t \in \mathbb{R}\}$ in the case of Hopf bifurcations are hyperbolic, i. e. 1 is a simple Floquet multiplier of the non-autonomous periodic linear o.d.e.

$$\dot{z}(t) = D_x f(x(t); \lambda_0) z(t)$$

which is obtained by linearizing (21) along $x^\star$. By Theorem 25.15 in [3] this condition ensures that there are neighbourhoods $U \subset \mathbb{R}^K$ of λ_0, $V \subset \mathbb{R}^m$ of γ_0

- for all $\lambda \in U$ there is a unique hyperbolic closed orbit $\gamma(\lambda) \subset V$ that has the same stability properties that γ_0 has
- and more important, the mapping

  $$U \to \mathbf{2}^{\mathbb{R}^m}, \lambda \mapsto \gamma(\lambda)$$

  is continuous with respect to the Hausdorff metric on $\mathbb{R}^m$

Note that this theorem works for parameters λ of arbitrary dimension whereas bifurcation theory gets extremely unhandy for more than two parameters. So if we can show the existence of a closed orbit by applying the Hopf theorem to one parameter only, the stability theorem in Amann shows that the appearance of a closed orbit is indeed locally stable under perturbation of all parameters. For another proof see [1].

So we have the following result on the qualitative behaviour of the system in the replication phase:

Lemma 3.9 *Under the assumptions of Theorem 3.3 with linearly independent $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ for every $\epsilon > 0$ there is a $t_0 = t_0(w(0), x(0)) > 0$ such that the x-component of the system (29) converges to a periodic solution $\tilde{x}$ with Hausdorff distance $d(\tilde{x}, x^\star) < \epsilon$.*

Proof: According to the remarks before the Lemma we need $\|W - W^\star\| < \delta$ for some $\delta > 0$ (because of the continuous dependence) to get periodic solutions with the same stability properties. Theorem 25.15 in [3] also gives a common set $V \in \mathbb{R}^m$ such that these orbits are attractive for all initial values

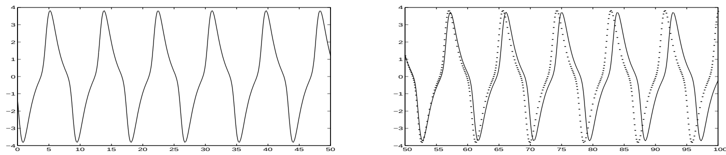

Figure 1: Output of the systems

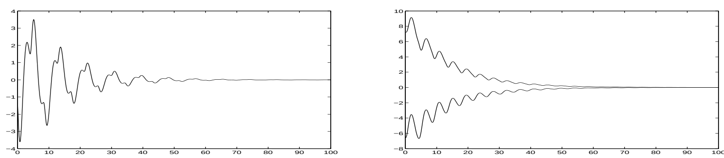

Figure 2: Tracking error and parameter error

$x \in V$. Because of exponential convergence according to Theorem 3.3 we have $\|W - W^\star\| < \delta$ and $\|x(t_0) - x^\star(t_0)\| < \delta$ for all t_0 sufficiently large. $\circ$

The convergence properties also give quantitative estimates on the approximation properties of x because of continuous dependence of solutions on the initial values like Theorem 3.16 in [14], also if $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ are not linearly independent.

4 Numerical Simulations

Let us consider a system with $n = 2$ where $\sigma(x) = \tanh(1.8x)$, $\tau = 1$, $\alpha = 0.1$, $w^\star{}_1 = -7.2$, $w^\star{}_2 = 6.6$. System (11) was started with $x(0) = 0$, $w(0) = 0$ and a point $x^\star(0)$ on the systems periodic orbit. The learning gains were $Q = I$. Fig. 4 shows the reference system output $y^\star$ on the left hand side and the output of the reference (dotted) and learning (solid) system when learning was stopped at $\bar{t} = 50$ with $w(t) = w(\bar{t})$ for all $t > \bar{t}$. The learning system produces an output very similar in shape but with greater period. In Fig. 4 both the tracking error $y - y^\star$ and the parameter error $w - w^\star$ clearly can be seen to converge exponentially. In another simulation we consider the above system in $n + 1 = 3$ dimensions as a system in 5 dimensions by trivially extending the reference weight vector to $w^\star = [0\,0\,-7.2\,6.6]^t$. Obviously this system can generate the same output as above. Due to the higher dimension convergence is much slower (Fig. 4, left). However, if we apply adatation of the gain matrix (25) we can achieve convergence in much shorter time (Fig. 4, right; $Q = 38I_4$). Note the different periods of time plotted in the two cases! One could argue that the adaptation of R simply increases the gain to speed up convergence. Indeed, $\|R^\star\|_2 \approx 38$ in our case. So this claim is refuted by the choice of Q. The convergence with $Q = 38I_4$ is not much faster than with $Q = 38I_4$. Also the results in (Fig. 4) using normalized gains from (28) confirm this even for $\eta = 1$ which has to be considered "large" in the averaging sense. This shows that it is necessary to treat different directions in parameter error space differently.

By the way we observe that it is not necessary to know the dimension of the reference system: An upper bound is enough. If convergence of some leading weights to zero occurs one may lower the dimension.

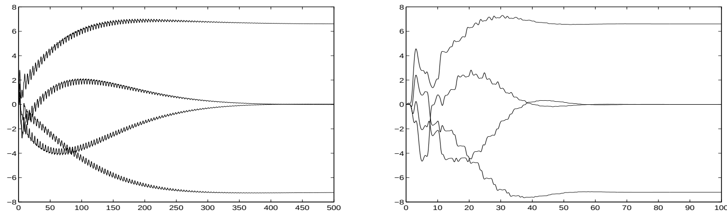

Figure 3: Parameter convergence with and without adaptation of gains

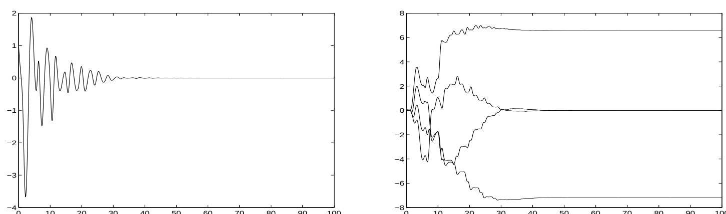

Figure 4: Parameter convergence with and without adaptation of gains

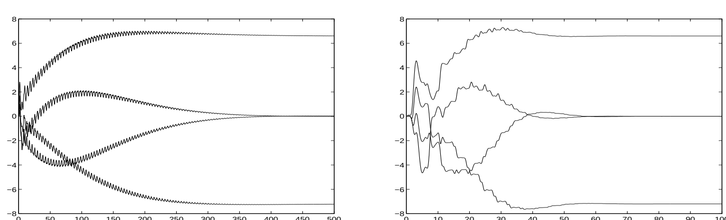

Figure 5: Parameter convergence with and without adaptation of gains

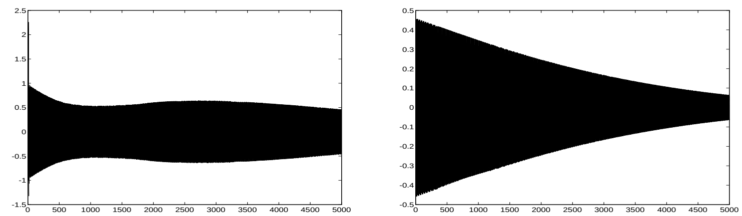

Figure 6: tracking error

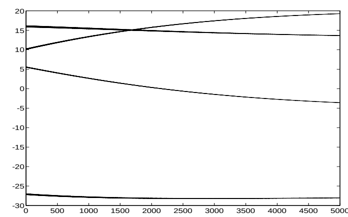

Figure 7: weight error

Fig. 4 and 4 show another example for an oscillatory system in 5d. The parameters are $w^{\star t} = [-4.68, 19.93, -27.56, 13.09]$, $\tau = 0.5$, $\alpha = 0.5$, $Q = I_4$.

5 Linear independence condition

We now investigate the LI condition of the convergence theorems. For the structure (7) we still hope to prove that this condition always holds. For more general structures like in Theorem 3.5 and Theorem 3.7 we will give a different approach in the next subsection.

Remark 5.1 *For the case $n = 2$, i. e. $x^\star(t) \in \mathbb{R}^3$, the learning rule will always converge, because of the periodicity of $x^\star$. This is an easy consequence of Lemma 2.3: Assume $a_1\sigma(x_1^\star) + a_2\sigma(x_2^\star) = 0$. No component of $x^\star$ is constant and therefore because of the bijectivity of σ neither $\sigma(x_1^\star)$ nor $\sigma(x_2^\star)$ is constant. This implies $a_1 \neq 0 \neq a_2$. Substituting $\sigma(x_2^\star) = \frac{a_1}{a_2}\sigma(x_1^\star)$ in the o.d.e. for $x_1^\star$ we get the contradiction that $x_1^\star$ is a nontrivial periodic solution of a one-dimensional autonomous o.d.e..*

Remark 5.2 (Stability of independence condition) *The remarks preceding Lemma 3.9 give that the linear independence of the componenents of a solution is also stable under perturbation of all parameters. Denoting the unit sphere in $\mathbb{R}^m$ by S^{m-1} we define functions*

$$\begin{aligned} G : \left\{X \in \mathbf{2}^{\mathbb{R}^m} : X\,compact\right\} \times S^{m-1} &\to \mathbb{R} \\ (X, a) &\mapsto \max_{x \in X} \left|a^t x\right| \\ g : \left\{X \in \mathbf{2}^{\mathbb{R}^m} : X\,compact\right\} &\to \mathbb{R} \\ X &\mapsto \min_{a \in S^{m-1}} G(X, a) \end{aligned}$$

$G(X, a)$ gives the maximum distance of a point in X from $a^\perp$, so $G(X, a) = 0$ iff $X \subset a^\perp$. Linear dependence of a compact set $X \subset \mathbb{R}^m$ is equivalent to $g(X) = 0$. One can easily verify that G and g are continuous w.r.t. Hausdorff metric. Since the mapping $\lambda \to \gamma_\lambda$ is continuous as well and

the existence of periodic solutions is stable under variations in λ as well by Lemma 3.9 we have continuity of $\lambda \to g(\gamma_\lambda)$. Therefore we get linearly independent functions $\sigma(x_1), \ldots, \sigma(x_n)$ for all parameters λ in an open environment of λ_0 of an asymptotically stable orbit.

Lemma 5.3 *Consider a periodic solution $x^\star$ of a m-dimensional network*

$$\dot{x} = -\tau x + W\sigma(x)$$

with $\det W \neq 0$. *Then the components $x_1^\star, \ldots, x_m^\star$ are linearly independent iff $\sigma(x_1^\star), \ldots, \sigma(x_m^\star)$ are linearly independent.*

Proof: Assume $a^t x = 0$ for some $0 \neq a \in \mathbb{R}^n$. Then

$$0 = \frac{d}{dt}(a^t x^\star) = -\tau a^t x^\star + a^t W\sigma(x^\star) = a^t W\sigma(x^\star)$$

and because of the invertibility $W^t a$ is a nonzero vector orthogonal to $\sigma(x^\star)$. Assume now that $b^t\sigma(x^\star) = 0$ for some $0 \neq b \in \mathbb{R}^n$. Define $a := W^{-t}b \neq 0$. Then

$$0 = \dot{(a^t x^\star)} = -\tau a^t x^\star + a^t W\sigma(x^\star) = -\tau a^t x^\star + b^t\sigma(x^\star) = -\tau a^t x^\star$$

Since $a^t x^\star$ is a periodic solution of the scalar o.d.e. $\dot{z} = -\tau z$ we have $a^t x^\star = 0$. ∘

Example 5.4 *An exponentially stable time-invariant linear system excited by a periodic input generates a periodic solution of the same period. Denote*

$$X_T = \{f : \mathbb{R} \to \mathbb{R} : f(t+T) = f(t) \forall t\}$$

Define the operators

$$\begin{aligned} P : X &\to X \\ (Px)(t) &:= \int_{-\infty}^{t} e^{-\tau(t-s)} x(s) ds \qquad (30) \\ P_\sigma : X &\to X \\ x &\mapsto P(\sigma \circ x) \qquad (31) \end{aligned}$$

It is easily seen that for $\alpha = 0$ one can construct systems of arbitrarily high dimension all generating the same output $y^\star$: Assume we are given a periodic solution $x^\star$ of a system (6) in dimension $n+1$. Add an equation

$$\dot{x}_0 = -\tau x_0 + \sigma(x_1)$$

and choose the weight $w_0 = 0$. This gives a system in dimension $n+2$ with the same structure of the weight matrix. For the right choice of the initial value $x_0(0)$ we get a stable periodic solution of the $n+2$-dimensional system whose last $n+1$ components are those of the $n+1$-dimensional one since x_0 does not affect the dynamics.
The definition of P_σ fits the single o.d.e.'s. Therefore $x_i = P_\sigma x_{i+1}$ for $i = 0, \ldots, n$ or $x_i = P_\sigma^{n+1-i} x_{n+1}$. On the other hand equation $n+1$ gives

$$x_{n+1} = \sum_{i=1}^{n} w_i P_\sigma x_i = \sum_{i=2}^{n} w_i x_{i-1}$$

We get the orthogonal vector $v = [w_1, \dots, w_n, 0, -1]^t = [w_1, \dots, w_n, w_0, -1]^t$ *We have* $span\{v\} = kernA(w)$. *Numerical simulations show, however, that* $\sigma(x_1), \dots, \sigma(x_n + 1)$ *are LI.*

Similar analysis shows that also for $\alpha \neq 0$ *for* $\det A(w) = 0$ *there is an invariant subspace* $kernA(w)$ *for the o.d.e.. Due to the special structure of* $A(w)$ *higher codimensions cannot occur.*

If we could show that for a periodic solution x *the LI of* $\sigma(x_1), \dots, \sigma(x_{n+1})$ *can only happen if* $span\{\sigma(x(t)) : t \in \mathbb{R}\} \neq \mathbb{R}^{n+1}$ *is an invariant subspace for the o.d.e. considered we would be done since such invariant spaces do not exist for* $\det A(w) \neq 0$ *(and this condition is generic).*

Also (affine) subspaces orthogonal to some unit vector cannont occur due to Lemma 2.3.

As this remark does not help for open loop identification we now give an approach that shows that the LI condition is generic (which has also been confirmed by all numerical simulations).

5.1 Persistent excitation

Instead of the LI of $\sigma(x_1^\star), \dots, \sigma(x_n^\star)$ let us study the stronger condition of LI of $\sigma(x_1^\star), \dots, \sigma(x_{n+1}^\star)$. By Lemma 5.3 this is equivalent to the LI of $x_1^\star, \dots, x_{n+1}^\star$ if $\det A(w) \neq 0$. However, the PE condition on the state $x^\star$ of (11) is equivalent to a PS condition on the state of a PLTV that can be constructed (like the error o.d.e. (16)) as follows:

By the mean value theorem we have for all $x \in \mathbb{R}$ that

$$\sigma(x) = \frac{\sigma(x) - \sigma(0)}{x - 0} x = \sigma'(\xi) x$$

for some $\xi = \xi(x)$ and the function $s : \mathbb{R} \to \mathbb{R}$

$$s(x) = \begin{cases} \sigma'(\xi(x)) & , x \neq 0 \\ \sigma'(0) & , x = 0 \end{cases}$$

defined by this is continuous because $\sigma \in C^1$.

$$\dot{x} = -\tau x + W \operatorname{diag}(\sigma'(\xi_i(t))) x + bu = -\tau x + W \operatorname{diag}(s(x^\star(t))) x + bu \tag{32}$$

has the same solution $x^\star$. For periodic $x^\star$ system (32) is periodic as well and under the assumption of hyperbolicity (which is generic in LTVs in the topology) $x^\star$ is the only periodic solution. We show that generically - in some sense to be specified - the state of a LTV is PS.

5.1.1 Persistent excitation in LTI

We recall some definitions concerning the PE property from [12], [11] and [15]. PE is a standard tool for adaptive control of linear systems. Its importance comes from the fact that it relates a property of a signal u, namely (33), to the stability of a linear time-varying system with special structure. As many results can be proven the same way for discrete time systems we also consider PE in linear difference equations.

Definition 5.5 (persistently exciting) *A function* $x \in L^\infty(\mathbb{R}^+, \mathbb{C}^n)$ *is called* persistently exciting (PE), *if there exist constants* T_0, δ_0, $\epsilon_1 > 0$ *and* $t_1 \in \mathbb{R}^+$ *so that for all* $t \geq t_1$ *there exists* $t_2 \in [t, t + T_0]$ *with*

$$\left| \frac{1}{T_0} \int_{t_2}^{t_2 + \delta_0} x(s)^* w ds \right| \geq \epsilon_1 \tag{33}$$

for all $w \in \mathbb{C}^n$ *with* $\|w\| = 1$.

This definition usually is made for real valued functions only but carries over to the comlex case. We need complex functions when we change to Floquet representation since even for real systems complex Floquet matrices may occur. PE of unbounded functions is not quite clear, see [12].

However, condition (33) is difficult to verify. For C^1 functions (in fact, even less smoothness is required, see [19]) one can replace (33) by: There exist $T_0, t_0, \epsilon_0 \geq 0$ such that

$$\int_t^{t+T_0} |x(s)^* w| \, ds \geq \epsilon_0 \tag{34}$$

for all $w \in \mathbb{C}^n$ with $\|w\| = 1$.

By further confining the class of signals to stationary and especially almost-periodic functions we get a much handier criteria.

Definition 5.6 (stationary function, autocovariance) *A function $u : \mathbb{R}^+ \to \mathbb{R}^n$ is called* stationary *iff the following limit, the* autocovariance Cov_u *of u, exists uniformly in s for all τ*

$$Cov_u(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_s^{s+T} u(t)u(t+\tau)^* dt \tag{35}$$

The most important class of stationary functions are almost periodic functions. Proofs for the following facts can be found in [5] or [9].

Remark 5.7 (Almost periodic functions) *For functions $f, g : \mathbb{R} \to \mathbb{C}$ define*

$$\langle f, g \rangle = \lim_{T\to\infty} \frac{1}{T} \int_\tau^{T+\tau} f(t)\bar{g}(t)dt \tag{36}$$

if the limit exists independently of τ. Obvoiusly $\langle u, u \rangle = Cov_u 0$ from Definition 5.6. This gives an scalar product on the vector space

$$\mathcal{A} = span\{f_\omega(\cdot) = \exp(i\omega\cdot) : \mathbb{R} \to \mathbb{C} : \omega \in \mathbb{R}\} \tag{37}$$

with the functions f_ω forming an orthogonal set. The closure of this space with respect to the norm induced by the above scalar product is denoted by $AP^c(\mathbb{R})$ which is a non-separable Hilbert space with complete orthogonal system given by the f_ω. Functions $f \in AP^c(\mathbb{R})$ are called almost periodic. *For $x \in AP(\mathbb{R})$ we denote by*

$$a(\omega; x) = \langle x, f_\omega \rangle \tag{38}$$

the Fourier coefficient *of the* Fourier exponent *or* frequency ω. *So x can be represented as a* Fourier series *which is an infinite sum of at most countably many elementary harmonic functions:*

$$x = \sum_{k\in\mathbb{N}} a(\omega_k; x) f_{\omega_k}, x = \sum_{k\in\mathbb{N}} a_k x f_{\omega_k}, \qquad \omega_k \in \mathbb{R} \tag{39}$$

with $\sum_k |a_k|^2 < \infty$. The subset

$$AP(\mathbb{R}) = \{x \in AP^c(\mathbb{R}) : x \text{ is continuous}\} \tag{40}$$

is called the space of continuous almost periodic functions. *Functions in $AP(\mathbb{R})$ are even uniformly continuous.*

Restrictions to (unbounded) subsets of $\mathbb{R}$ like $AP(\mathbb{R}^+)$ and higher dimensional analoga like $AP(\mathbb{R}^n)$ are defined as usual with Fourier coefficients calculated componentwise.

One can also define almost periodic functions in discrete time starting with the set

$$\tilde{\mathcal{A}} = \{\exp(i\omega\cdot) : \mathbb{Z} \to \mathbb{C}; \omega \in (-\pi, \pi]\} \tag{41}$$

and the scalar product

$$\langle f, g\rangle = \lim_{T\to\infty} \frac{1}{T} \sum_{t=\tau}^{T+\tau} f(t)\bar{g}(t) \tag{42}$$

Then analogous properties hold for the spaces $AP^c(\mathbb{Z})$ and $AP(\mathbb{Z})$.

The fundamental results establishing the basic connections between the notions of stationarity, almost periodicity and persistent excitation are:

Lemma 5.8 *[6] Let $u : \mathbb{R}^+ \to \mathbb{C}^n$ be stationary and $v : \mathbb{R}^+ \to \mathbb{R}^n$ in $L^2(\mathbb{R}^+, \mathbb{C}^n)$. Then:*

1. *The function u is persistently exciting iff $Cov_u(0) > 0$.*
2. *The functions u and $u+v$ have the same autocovariance functions: $Cov_u = Cov_{u+v}$. Therefore u is persistently exiting iff $u + v$ is.*

Lemma 5.9 *Let $x \in AP(\mathbb{R}, \mathbb{C}^n)$ ($x \in AP(\mathbb{Z}, \mathbb{C}^n)$). Then*

$$span\{x(t) : t \in \mathcal{T}\} = span\{a(\omega; x) : \omega \in \mathbb{R}\} = im(Cov_x(0))$$

Especially, the following are equivalent:

1. *The component functions $x_1, \ldots, x_n$ are linearly independent.*
2. *There exist frequencies $\omega_1, \ldots, \omega_n$ such that the amplitudes $a_k = a(\omega_k; x)$, $k = 1, \ldots, n$ are linearly independent.*
3. *$Cov_x(0) > 0$*

Proof: Of course $\text{span}\{x(t) : t \in \mathcal{T}\} \subset \text{span}\{a(\omega; x) : \omega \in \mathbb{R}\}$ due to Fourier representation for continuous and discrete time. Now first consider discrete time. By the Fourier inversion theorem for almost periodic functions there is a bijective linear correspondence between $(x_t)_{t\in\mathbb{Z}}$ and the amplitudes $(a_\omega)_{\omega\in[-\pi,\pi)}$ (of which only countably many are non-zero). Because of bijectivity $\dim \text{span}\{x_t : t \in \mathbb{Z}\} = \dim \text{span}\{a_\omega : \omega \in [-\pi, \pi)\}$ (we need no closure since $\dim \mathbb{C}^n < \infty$). Since we already have one inclusion and $\mathbb{C}^n$ is finite dimensional we get equality.

Now for the second inclusion in continuous time, with a different idea: Assume $a(\omega; x) \neq 0$ but $a(\omega; x) \notin \text{span}\{x(t) : t \in \mathcal{T}\}$. This obviously contradicts the definition of $a(\omega; x)$.

Now 1. and 2. are obviously equivalent 2. and 3. are equivalent as by the orthogonality of the harmonics and the definition of autocovariance and scalar product

$$\text{im}(\text{Cov}_x(0)) = \sum_{\omega : a(\omega;x)\neq 0} a(\omega; x)a(\omega; x)^*$$

$\circ$

Theorem 5.10 *([12] Lemma 6.6, Theorem 6.3, [15], Theorem 2.7.2.) Consider a stable LTI with input $u \in AP^c(\mathcal{T})$.*

1. *If (A,b) is controllable then the steady state response x is PS iff the u contains at least n frequencies.*

2. *If (A,b) is not controllable then x is not PS.*

Proof:

1. See the proof for the LTV case, Remark 5.12.

2. The standard proof is [15], Theorem 2.7.2. based on the transformation to controllable canonical form. We give an alternative proof based on another transformation in state space that resembles our approach to LTVs: Assume A in Jordan canonical form $A = \mathrm{diag}_{k=1}^{K}(J_{n_k}(\lambda_k))$, $n = \sum_{k=1}^{K} n_k$ and $b = (b^{(1)} \ldots, b^{(K)})$, $b^{(k)} \in \mathbb{C}^{n_k}$. Controllability implies $b_{n_k}^{(k)} \neq 0$. As Jordan blocks commute with Toeplitz matrices we can further get $b^{(k)} = e_{n_k}$ (and all matrices of this kind are controllable as one can check, so-called confluent matrices).

   By Lemma 5.8 PS of x is equivalent to the existence of frequencies $\omega_1, \ldots, \omega_n$ such that $\mathrm{span}\{a(\omega_k; x) : k = 1, \ldots, n\} = \mathbb{C}^n$. Choose any $\omega_1, \ldots, \omega_n$ with $a(\omega_k; u) \neq 0$ whose existence is guaranteed by the assumption. The transfer function of a Jordan block $J_m(\mu)$ is $(\tilde{s}I - J_m(\mu))^{-1} =: M$ with $M_{rs} = -(\frac{-1}{\mu - \tilde{s}})^{s-r+1}$ for $r \leq s$ and $M_{rs} = 0$ otherwise. The amplitude vector for ω_k then takes the form

$$
\begin{aligned}
a_k \quad = \quad & -[(\lambda_1 - i\omega_k)^{-n_1}, \ldots, (\lambda_1 - i\omega_k)^{-1}, \\
& (\lambda_2 - i\omega_k)^{-n_2}, \ldots, (\lambda_2 - i\omega_k)^{-1}, \\
& \ldots \\
& (\lambda_K - i\omega_k)^{-n_K}, \ldots, (\lambda_2 - i\omega_k)^{-1}]
\end{aligned}
$$

   Multiplying these vectors $\tilde{a}_k := (\neq \prod_{l=1}^{K}(i\omega_k - \lambda_l))a_k$ respectively to get rid of the denominators the linear independence of the vectors $\tilde{a}_k$ or a_k, $k = 1, \ldots, n$ is equivalent to the unique interpolation property of the polynomials

$$
p_{km}(s) = (s - \lambda_k)^m \prod_{k \neq l} (s - \lambda_l)^{n_l}
$$

   $k = 1, \ldots, K$, $m = 0, \ldots, n_k - 1$ for the interpolation points $i\omega_k$. This problem, again, is called to Hermite interpolation and well known to posess the unique interpolation property (In the case of diagonalizable matrix A we get the special case of Lagrange interpolation).

$\circ$

Considering the state of an exponentially stable system as a function on a semi-infinite intervall (a, ∞) we have PE for arbitrary initial values $x(a)$ since the difference to the steady state response is a L^2-function that does not matter according to reftLemmalemStatPE. Considering the steady state solution one may relax the assumption of stability of A in Theorem 5.10 to hyperbolicity.

Theorem 5.11 *Let $A(t) \in \mathbb{R}^{m \times m}$ be an matrix of L^∞ functions such that the o.d.e. $\dot{z} = A(t)z$ is uniformly asymptotically stable. Let $U(t) \in \mathbb{R}^{n \times m}$ be an matrix of L^∞ functions. Let $P(t) \in \mathbb{R}^{m \times m}$ be a symmetric positive definite matrix of bounded absolutely continuous functions with $\dot{P} \in L^\infty$ such that $\dot{P} + A^t P + PA$ is uniformly negative definite. Let $Q(t) \in \mathbb{R}^{n \times n}$ be an symmetric positive definite matrix of L^∞ functions with*

$$
a_1 I \leq Q(t) \leq a_2 I
$$

for some constants $0 < a_1 \leq a_2 < \infty$. Then the system

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} A & U^t \\ -QUP & 0 \end{bmatrix} \tag{43}$$

is uniformly asymptotically stable iff U is persistently exciting.

Proof: Note that the uniform negative definiteness of $A+A^t$ by Krasovskii's theorem ([12], Theorem 2.11) ensures the existence of matrix functions P with $\dot{P} + A^tP + PA = -S$ for uniformly positive definite matrix functions $S \in L^\infty$. For $Q = I$ the theorem is proven in [10], the extension to the case of arbitrary constant Q is Corollary 3.3 in [14] and carries over to functions $Q(t)$ uniformly bounded from above and below with the same proof. ∘

An important point is that a signal gives a PE steady state response for all controllable LTI in a fixed dimension or for none. For almost periodic signals this follows from Theorem 5.10 and also holds for arbitrary bounded signals, see [12], Theorem 6.2. Signals resulting in a PE state for controllable systems of dimension n are sometimes called *persistently exciting* whereas the state itself is called *persistently spanning (PS)* instead.

5.1.2 Persistent excitation in PLTV

We now generalize results on the PE property to PLTV.

Remark 5.12 *One might suspect that Theorem 5.10 also holds for LTV under some controllability assumptions. For exponentially stable systems with matrices in Kalman decomposition form (that can always be constructed for so-called constant rank systems, see [16]*

$$\dot{x} = \dot{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \end{bmatrix} u, \qquad x_1 \in \mathbb{C}^{n_1}, x_2 \in \mathbb{C}^{n_2} \tag{44}$$

- or the discrete time analogue - and $u \in AP(\mathcal{T})$ at least one part of Theorem 5.10 holds: The x_2-component converges to 0 exponentially and therefore is an L^2-function on $\mathbb{R}^+$ (or $\mathbb{N}$). So $Cov_{x_2}(0) = 0$ and x is not PS. Note that the x_1-component need not be controllable in this case but the whole system is not controllable in any sense provided $n_2 > 0$.

However for LTVs the transformation to the above form in general is time-varying and may change the PS property. For any $x : \mathbb{R} \to \mathbb{C}^n$ we can find a unitary matrix function as smooth as x $S : \mathbb{R} \to \mathbb{C}^{n \times n}$ such that $S(t)x(t) \in span\{v\}$ for some constant vector $v \in \mathbb{C}^n$. So Sx is contained in a one-dimensinal subspace and not PS by Lemma 5.9. Conversely even a constant function $x(t) \equiv x_0 \neq 0$ can be transformed into a PS function using suitable $S(t)$ (e.g. if $v = e_1$ choose S such that $S(t)e_1 = [\exp(i\omega t), \exp(i2\omega t), \ldots, \exp(in\omega t)]^t$, $\omega \neq 0$).

A consequence is that for LTV the PE of a signal can only be defined with respect to a given system and a given (possibly time-varying) choice of coordinates whereas in LTI a signal is PE for all stable, controllable systems or none (for almost periodic functions this is a consequence of Theorem 5.10; for more general functions see). Anyway, in both cases a time-invariant change of coordinates does not affect the PE properties of a signal for a *given* system.

The following results will show that for LTV

$$\dot{x} = A(t)x + b(t)u, \qquad x(t) \in \mathbb{C}^n \tag{45}$$

we can only define PE like this:

Definition 5.13 *Consider an exponentially stable system (45). A function* $u \in L^\infty(\mathbb{R}+, \mathbb{C})$ *is called* persistently exciting for (45) *if the steady state response fulfils (33).*

There are several definitions for controllability of LTVs. The nicest properties hold for so-called *constant rank systems*, see [16]. The most important representatives of this class are systems with analytic A and b. In our "linearization" method we get analyticity if we make the

Assumption 5.14 *The sigmoidal function* σ *is real analytic:* $\sigma \in C^\omega(\mathbb{R})$.

In practice this is no serious restriction since the main examples in neural networks, $\sigma = \tanh$ and $\sigma = \arctan$, are real analytic.

With the results in the LTI case in mind, consider now some examples of systems excited by superposition of harmonics:

Example 5.15 *One can show by examples that controllability of a LTV combined with n harmonics in the input is not sufficient for the PE of the steady state. We construct such a case for* $n = 2$ *by choosing constant* $A = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$, *fixing a basic frequency* $\omega = 1$ *and time-varying gains* $b(t) = c + \exp(i\omega t)d$ *with* $u = \exp(i\omega t) + \exp(i2\omega t)$. *Choosing* $c = [1, 2]^t$, $d = [2 + i, 2.8 + 1.6i]^t$ *one gets a steady state solution which lies in a one-dimensional subspace of* $\mathbb{C}^2$ *as simple calculations show (For numerical verification make sure that your code can handle complex o.d.e.'s). We do prove this at the moment and refer to Lemma 5.9 where the conditions will show to construct such examples. Calculating the time-varying observability matrix one also sees that the system is controllable!*

So we need another condition on the amplitudes of the harmonics.

On the other hand, given a LTV and frequencies $\omega_1, \ldots, \omega_n$ *one can suspect from this example that the amplitudes* $(u_1, \ldots, u_n)$ *resulting in non-PS state always are contained in a set that is the union of finitely many proper subspaces of* $\mathbb{C}^n$ *so that n frequencies still are sufficient for PE for a generic set of amplitudes.*

Example 5.16 *On the other hand even a constant nonzero input can generate PS state like in*

$$\dot{x} = diag(\lambda_l)x + (\sum_{k=1}^{n} \exp(i\omega_k t)b_k)u$$

when the $\lambda_l \in \mathbb{C}$ *and* $b_k \in \mathbb{C}^n$ *are chosen in such a way that* $span\{\frac{1}{i\omega_k - \lambda}b_k : k = 1, \ldots, n\} = \mathbb{C}^n$, *e. g.* $\lambda = \lambda_l$ *for all l and linearly independent* b_k.

Example 5.17 *A single harmonic* $b(t) = \exp(i\omega t)b_\omega$, $b_\omega \in \mathbb{C}^n$ *in the gains* $b(t)$ *reduces to the LTI case (which is the special case* $\omega = 0$*) when* A *is not time-varying since*

$$\dot{x} = Ax + (\exp(i\omega t)b_\omega)(\sum_{k=1}^{n} u_k \exp(i\omega_k t) = Ax + b_\omega(\sum_{k=1}^{n} u_k \exp(i(\omega + \omega_k)t)$$

So one has to consider the frequency content in b and u. First we give controllability results for LTVs with constant $A(t)$ in (45): Consider the time-varying controllability matrix $B(t) = [b_0(t), b_1(t), \ldots]$ with

$$b_0(t) := b(t), \qquad b_{i+1}(t) = A(t)b_i(t) - \frac{d}{dt}b_i(t)$$

(see [17]). Recall that a C^∞ system (45) is *controllable on an intervall* $[t_0, t_1]$ (in the sense that we have $\mathcal{R}(t_0, t_1) = \mathbb{C}^n$ where the *reachability space* is defined as

$$\mathcal{R}(t_0, t_1) = \left\{ \int_{t_0}^{t_1} \Phi(t,s) b(s) u(s) ds : u \in L^\infty[t_0, t_1] \right\}$$

) if rank$B(\tilde{t}) = n$ for some $\tilde{t} \in [t_0, t_1]$. For C^ω systems rank$B(\tilde{t}) = n$ implies controllability on $[t_0, t_1]$. A system is called *controllable* if it is controllable on every nontrivial interval.

Lemma 5.18 *Consider (45) with A constant in Jordan canonical form $A = diag_{k=1}^K(J_{n_k}(\lambda_k))$, $n = \sum_{k=1}^K n_k$ and $b = (b^{(1)} \ldots, b^{(K)}) \in C^\infty$, $b^{(k)} : \mathbb{R} \to \mathbb{C}^{n_k}$. Set $\mathcal{I}(\lambda) = \{k : \lambda_k = \lambda\}$, $\lambda \in spec(A)$. Denote the last component function of $b^{(k)}$ by $\tilde{b}_k := b_{n_k}^{(k)}$, $k = 1, \ldots, K$. If (45) is controllable on $[t_0, t_1]$ the sets $\mathcal{B}(\lambda) := \left\{ \tilde{b}_k : k \in \mathcal{I}(\lambda) \right\}$, $\lambda \in spec(A)$ are linearly independent. If $b \in C^\omega$ and the system is controllable on any nontrivial intervall then $\mathcal{B}(\lambda)$ are linearly independent on any intervall.*

Proof: As A is constant the controllability matrix consists of

$$\begin{aligned} b_0 &= b \\ b_1 &= Ab - \dot{b} \\ b_2 &= A^2 b - 2A\dot{b} + \ddot{b} \qquad (46) \\ \ldots & \qquad (47) \end{aligned}$$

By the structure of the Jordan blocks only terms involving $\tilde{b}_k$ and its derivatives appear in the last component of each block.

Assume that $\mathcal{B}(\lambda)$ is linearly dependent, $\sum_{k \in \mathcal{I}(\lambda)} \beta_k b_{n_k}^{(k)}$. Then also $\mathcal{B}^{(n)}(\lambda) := \left\{ \frac{d^n}{dt^n} \tilde{b}_k : k \in \mathcal{I} \right\}$ are linearly dependent. By the structure of the controllability matrix where only derivatives of these functions occur in the components corresponding to the last rows of the Jordan blocks we can choose the vector $v \in \mathbb{C}^n$ with entries β_k in position corresponding to $\tilde{b}_k$ satisfying $v^t B(t) = 0$ for all $t \in [t_0, t_1]$. So the system is not controllable on any intervall since $\mathcal{R}(t_0, t_1) \subset v^\perp$ for all t_0, t_1. Therefore controllability implies linear independence of all $\mathcal{B}(\lambda)$.

Assume now that $b \in C^\omega$ and independence condition on the $\mathcal{B}(\lambda)$ holds. For each of the functions $\tilde{b}_k$ the set of zeros is at most countable having no limit point since otherwise analyticity would give $\tilde{b}_k = 0$ contradicting controllability. So there exists t such that $b_k(t) \neq 0$ for all k. Fix t and assume that rank$B(t) < n$. Choose a constant coordinate transform with upper triangular (time-independent) Toeplitz matrices such that the transformed vector $b(t)$ has the same structure as the constant gain vector b in the proof of Theorem 5.27, i. e. $b^{(k)}(t) = e_{n_k}$. By the triangular structure of the transformation the last components in each block still contain only the functions $\tilde{b}_k$ multiplied by a non-zero scalar so the independence assumption still hold. (The special form with unit vectors $b_k(t)$ holds at time t, not necessarily on all $\mathbb{R}$.) Choose $0 \neq v \in \mathbb{C}^n$ such that $v^t B(t) = 0$ by the assumption that rank$B(t) < n$, i. e. $v^t b_m(t) = 0$ for all $m \in \mathbb{N}_0$. The structure (46) and analyticity give $v^t b_m \equiv 0$ for all m. Using this again in (46) we get inductively $v^t A^m b = 0$ for all $m \in \mathbb{N}_0$.

Now consider again time t. Here the matrix

$$[b(t), Ab(t), A^2 b(t), \ldots] = \begin{bmatrix} A_1(t) \\ \vdots \\ A_K(t) \end{bmatrix}$$

consists of confluent-like blocks

$$A_k(t) = \begin{bmatrix} & & & \ddots & \\ & & 1 & 3\lambda_k & \dots \\ & 1 & 2\lambda_k & 3\lambda_k^2 & \dots \\ 1 & \lambda_k & \lambda_k^2 & \lambda_k^3 & \dots \end{bmatrix} \in \mathbb{C}^{n_k \times n}$$

that do not give full rank because of linear dependencies for A_k belonging to the same eigenvalue. Considering however only the last, the n_k-th rows of these matrices (since at t the remaining components are zero), one may write the linear dependence as the following sum of row vectors in $\mathbb{C}^n$

$$0 = \sum_{\lambda \in \mathrm{spec}(A)} \big(\sum_{k \in \mathcal{B}(\lambda)} \tilde{v}_k ({e_{n_k}}^t A_k(t)) \big)$$

where the $\tilde{v}_{k,\lambda}$ are the components of v belonging to $\tilde{b}_k$. Denote $S := |\mathrm{spec}(A)|$. Since the last rows of the $A_k(t)$ corresponding to the same Eigenvalue λ are the same we have

$$0 = \sum_{\lambda \in \mathrm{spec}(A)} \big(\sum_{k \in \mathcal{B}(\lambda)} \tilde{v}_k \big) ({e_{n_k}}^t A_k(t))$$

Considering the first S columns of this equation one gets a Vandermonde matrix since the Eigenvalues have been grouped. So we have

$$\sum_{k \in \mathcal{B}(\lambda)} \tilde{v}_k = 0$$

for all λ.

Now look at the first column of $v^t B \equiv 0$ which holds on $\mathbb{R}$. Using the last equation we get for all λ

$$\sum_{k \in \mathcal{B}(\lambda)} \tilde{v}_k \tilde{b}_k = 0$$

By the linear independence of the $\tilde{b}_k$ we finally get that $v = 0$. $\circ$

Corollary 5.19 *Under the assumptions and notations of (5.18), let in addition $b \in AP(\mathbb{R}, \mathbb{C}^n)$. The the system is controllable iff the for all $\lambda \in spec(A)$*

$$span\{(a(\omega; \tilde{b}_k))_{k \in \mathcal{I}(\lambda)} : \omega \in \mathbb{R}\} = \mathbb{C}^{|\mathcal{I}(\lambda)|}$$

Proof: Lemma 5.18, Lemma 5.9 $\circ$

Example 5.20 *Choosing the transformation $x(t) = \Phi(t, t_0) z(t)$ we can transform (45) into $\dot{z} = \tilde{b}(t) u$ where controllability corresponds to LI of $\tilde{b}_1, \dots, \tilde{b}_n$ as also can be checked directly. However the transformation is not a Lyapunov transformation for exponentially stable systems.*

The situation of the Lemma can always be achieved via a Lyapunov tranformation for periodic constant rank systems.

We now restrict to analytic LTVs with periodic A and b: $A \in \mathcal{P}_T(\mathbb{C}^{n \times n})$, $b \in \mathcal{P}_T(\mathbb{C}^n)$R, $u \in \mathcal{P}_T(mC)$; that is, the special case of almost periodic functions whose module (see [9] for a definition) since this is the case in our neural network application. Analyticity gets rid of functions not identically zero but zero on some intervall. The results can be generalized to general modules with the same idea but notation gets very unhandy whereas periodic functions take us to the realm of ordinary Fourier series.

For the multiplication of Fourier we need a result on convolutions:

Lemma 5.21 *Let $P \in l^1(\mathbb{C}^{p\times n})$, $Q \in l^1(\mathbb{C}^{n\times q})$ be the Fourier series of analytic functions. If $span\{P_t e_i : t \in \mathbb{Z}\} = \mathbb{C}^{p\times n}$ then there is an open and dense subset $\mathcal{Q} \subset l^1(\mathbb{C}^{n\times q})$. such that*

$$span\{(P * Q)_t : t \in \mathbb{Z}\} = \mathbb{C}^{p\times q} \tag{48}$$

for all $Q \in \mathcal{Q}$. (48) also holds if $span\{Q_t : t \in \mathbb{Z}\} = \mathbb{C}^{n\times q}$ for all P in an open and dense subset $\mathcal{P} \subset l^1(\mathbb{C}^{p\times n})$.

Proof: We only prove the first statement. For $p = n = q = 1$ the case is clear since there are no zero divisors for the convolution due to analyticity and $P \neq 0$.

For $p = q = 1$, n arbitrary, observe that the mapping $l^1(\mathbb{C}^{n\times 1}) \to l^1(\mathbb{C})$, $Q \mapsto P * Q$ is linear and continuous and not identically zero. So the kernel is a proper subspace, i. e. its complement a open and dense set.

Considering scalar products $v^t P * Q$ with vectors $0 \neq v \in \mathbb{C}$ we get the result. ∘

Proposition 5.22 (PE for PLTV in canonical form) *Consider a controllable exponentially stable C^ω PLTV (45) with A and b partitioned as in Lemma 5.18. Let $b \in AP(\mathbb{R}, \mathbb{C}^n)$, $b(t) = \sum_{k\in\mathbb{Z}} b_k \exp(ik\omega t)$ and $u \in AP(\mathbb{R}, \mathbb{C})$, $u(t) = \sum_{l\in\mathbb{Z}} u_l \exp(il\omega t)$ where $K + L \geq n$. Then there exists an open and dense set $\mathcal{U} \subset \mathbb{C}^{\mathbb{Z}}$ such that the steady state is PS if $u = (u_l)_{l\in\mathbb{Z}} \in \mathcal{U}$.*

Proof: Using the notation of Lemma 5.18 and Lemma 5.19 we get for all $\lambda \in \operatorname{spec}(A)$ that the matrices (with columns indexed by $\mathbb{R}$)

$$\hat{A}(\lambda) = \operatorname{span}\{(a(\omega; \tilde{b}_k))_{k\in\mathcal{I}(\lambda)} : \omega \in \mathbb{R}\} = \mathbb{C}^{|\mathcal{I}(\lambda)|}$$

By Lemma 5.21 we get an generic set $\mathcal{U}$ such that also bu fulfils this condition instead of b.

Now the same arguments using confluent-like matrices as in the proof of Lemma 5.18 give the result; the only difference being that in the last step of the proof we now have an independence condition on the rows of the matrix $\hat{A}(\lambda)$ built from Fourier coefficients instead of the time-domain functions $\tilde{b}_k$ (actually, the rows of $\hat{A}(\lambda)$ are the Fourier transform of the $\tilde{b}_k$; one might also appeal to Lemma 5.9). ∘

Corollary 5.23 *Consider a controllable exponentially stable C^ω LTV (45) with A and b partitioned as in Lemma 5.18. Let $b \in AP(\mathbb{R}, \mathbb{C}^n)$, $b(t) = \sum_{k=1}^{K} b_k \exp(i\omega_k t)$ and $u \in AP(\mathbb{R}, \mathbb{C})$, $u(t) = \sum_{l=1}^{L} u_l \exp(i\omega_l t)$ where $K + L \geq n$. Then there exists an open and dense set $\mathcal{U} \subset \mathbb{C}^K$ such that the steady state is PS if $u = (u_1, \dots, u_K) \in \mathcal{U}$.*

Proof: If the support of the Fourier series of b and u has K resp. L points the support of the Fourier series of $b * u$ has at least $K + L \geq n$ points. Under the controllability assumptions one gets that the Fouries amplitudes span $\mathbb{C}^n$ generically as in Proposition 5.22. Again Lemma 5.9 transports the span condition to the steady state. ∘

Remark 5.24 *For almost periodic functions with modules whose integral base consists of K frequencies for b and L frequencies for u one may show that generically the steady state is PS if $KL \geq n$.*

It remains to study the effect of coordinate transformations. By Example 5.15 and Example 5.15 the PS property may be created or constructed. We now restrict to PLTV: Let $A \in \mathcal{P}_T(\mathbb{C}^{n\times n})$, $b \in \mathcal{P}_T(\mathbb{C}^n)$. By Floquet theory there exist $P \in \mathcal{P}_T(\mathbb{C}^{n\times n})$, $C \in \mathbb{C}^{n\times n}$ such that $\Phi(t, 0) = P(t)\exp(tC)$ changing the equation into

$$\dot{z} = Cz + P(t)^{-1} b(t) u$$

where $P(t)z(t) = x(t)$.

Lemma 5.25 *Let $P \in \mathcal{P}_T(\mathbb{R}, GL(n,\mathbb{C}))$, $x \in \mathcal{P}_T(\mathbb{R},\mathbb{C}^n)$, $P, x \in C^\omega$. Assume span$\{x_t : t \in \mathbb{R}\} = \mathbb{C}^n$ and denote the Fourier coefficients of P and x by $\hat{P}$, $\hat{x}$. Then there is a generic set $\mathcal{P} \subset l^1(\mathbb{C}^{n\times n})$ such that span$\{(Px)_t : t \in \mathbb{R}\} = \mathbb{C}^n$ if $\hat{P} \subset \mathcal{P}$.*

Proof: The Fourier coefficients of Px are $\hat{P} * \hat{x}$, so apply Lemma 5.21 and Lemma 5.9 ∘

Remark 5.26 *One may also consider continuous periodic transformations P and argue that whenever span$\{(Px)_t : t \in \mathbb{R}\} = \mathbb{C}^n$ also transformations $P + \Delta P$ with ΔP periodic, continuous, $\|\Delta P\|_{\infty,[0,T]}$ small enough give span$\{((P+\Delta P)x)_t : t \in \mathbb{R}\} = \mathbb{C}^n$ using continuous dependence of the transformed periodic orbit on ΔP in the Hausdorff metric as in the proof of Lemma 3.9.*

Theorem 5.27 (PE for PLTV) *Consider a analytic PLTV (45) with analytic almost periodic input u and Floquet representation $\Psi(t) = P(t)\exp(t\Gamma)$. Then for generic P, u in the sense of Proposition 5.22 and Lemma 5.25 the steady state of the system is PS.*

Proof: Proposition 5.22, Lemma 5.25 ∘

6 Discrete Time

An analogous learning rule can be given for the discrete time counterpart of system (11):

$$\begin{aligned} x^\star(t+1) &= A(w^\star)\sigma(x^\star(t)) + by^\star(t)(A(w^\star) + bc^t)\sigma(x^\star(t)) \\ y^\star(t) &= c^t x^\star(t) \\ x(t+1) &= A(w)x(t) + b\sigma(y^\star(t)) \quad (49) \\ y(t) &= c^t x(t) \\ w(t+1) &= w(t) + f(w, y^\star, x) \quad (50) \end{aligned}$$

The analogue of [14] works again:

$$w(t+1) = w(t) - \epsilon(y(t) - y^\star(t))\sigma(\underline{x}(t)) \quad (51)$$

Theorem 6.1 *Consider (11) with $\sigma \in C^1$ satisfying (4). Let $x^\star$ be nontrivially almost periodic. Let $x(0)$ and $w(0)$ be given. Then there exists a $\epsilon_0 > 0$ such that for all $0 < \epsilon < \epsilon_0$: The learning rule (51) solves the output tracking problem (12) and also the state error converges to 0. If $\sigma(x_1^\star), \ldots, \sigma(x_n^\star)$ are linearly independent which is equivalent to the PE property of $\sigma(\underline{x^\star})$ then the state and parameter error converge to zero exponentially. Otherwise the weight vector converges to an affine subspace containing $w^\star$:*

$$w(t) \rightarrow w^\star + \{\sigma(\underline{x^\star}(t)) : t \in \mathbb{R}\}^\perp \quad (52)$$

Proof: The proof proceeds along the lines of [8] using a discrete time version of the partial convergence theorem for (52). Exponential convergence under the PE condition is proven using the averaging theory developed in [14]. Details will be published elsewhere. ∘

The proof uses the discrete time analogue of (45):

$$x(t+1) = A(t)x(t) + b(t)u(t), \qquad x(t) \in \mathbb{C}^n \quad (53)$$

Questions concerning the PS property can be answered as in discrete time under the additional assumption of invertibility of the system matrices of the corresponding LTV since we can define a Floquet-like transformation to time-independent system matrix only in this case:

Proposition 6.2 (Floquet theorem for discrete time) *Consider a discrete time linear difference equation*

$$x(t+1) = A(t)x(t) \tag{54}$$

with $A \in \mathcal{P}_T$ *for some* $T \in \mathbb{N}$ *and* $A(t) \in GL(n, \mathbb{C})$ *for all* t. *Then the fundamental solution* $\Psi(t) = \Phi(t, 0)$ *can be represented as* $\Psi(t) = P(t)\Gamma^t$ *for some* $P \in \mathcal{P}_T$, $\Gamma \in GL(n, \mathbb{C})$.

Proof: By periodicity and uniqueness of solution we have $\Psi(t+T) = \Psi(t)C$ for some $C \in \mathrm{GL}(n, \mathbb{C})$. Write $C = \Gamma^T = \exp T\tilde{\Gamma}$ for some $\Gamma \in \mathrm{GL}(n, \mathbb{C})$ as in continuous Floquet theory and set $P(t) := \Psi(t)\Gamma^{-t}$. We have $P \in \mathcal{P}_T$ since

$$P(t+T) \quad = \quad \Psi(t+k)\Gamma^{-(t+T)} = \Psi(t)\Gamma^{-T}\Gamma^{-t} = \Psi(t)\Gamma^{-t} = P(t)$$

and by definition the representation $\Gamma \in \mathrm{GL}(n, \mathbb{C})$. ∘

If there are only Jordan blocks of size 1 for the zero Eigenvalue in $\Psi(T)$ one may define a Floquet-like representation under further assumptions. For nilpotent blocks N_m of order $m \geq 2$ the equation $N_m = X^T$ has no solution.

Remark 6.3 *The results on the number of harmonics necessary to achieve PS (stated and proven analogously to Corollary 5.23) are especially intuitive in discrete time: A* T*-periodic function may only contain* T *harmonics which corresponds to* $\dim span\{x_t : t = 1, \ldots, T\} \leq T$.

7 Conclusions

For a special type of network we have presented an simple and fast learning rule working with neural network dynamics only. The condition of linear independence remains to be investigated. Similar results can be obtained for discrete time systems with more technical arguments. The full weight matrix can be identified using an open loop version of the algorithm under the quite restrictive assumption $\|W^\star\|_2 \|\sigma'\|_\infty < \tau$ when all of the state is accessible. The algorithm heavily relies on the fact that the parameters enter the network dynamics linearly and is therefore not easily extended to more general systems as those in [2] that are dense in many spaces.

Acknowledgement

This work was supported by Deutsche Forschungsgemeinschaft and Graduiertenkolleg Technomathematik.

Numerical results were obtained on computers of the RHRK, Universität Kaiserslautern.

A Notation and Abbreviations

l^p	space of p-summable functions on $\mathbb{Z}$
L^p	space of Lebesgue p-integrable functions on $\mathbb{R}$
LI	linearly independent
LTI	linear timeinvariant system
LTV	linear timevariant system
PD(n)	positive definite symmetric matrices
$\mathcal{P}_T$	space of functions with period T
PE	persistently exciting
PLTV	periodic linear timevariant system
PS	persistently spanning
PSD(n)	positive semidefinite symmetric matrices
spec(A)	spectrum of A

References

[1] R. Abraham and J. Robbin. *Transversal Mappings and Flows.* W. A. Benjamin, Inc., New York, 1967.

[2] F. Albertini and E. D. Sontag. For neural networks, function determines form. *Neural Networks*, 7:975–990, 1993.

[3] H. Amann. *Gewöhnliche Differentialgleichungen.* de Gruyter, Berlin, 1983.

[4] A. F. Atiya. Learning on a general network. In *Neural Information Processing Systems*, pages 22–30. American Institute of Physics, 1988.

[5] H. Bohr. *Almost Periodic Functions.* Chelsea, New York, 1974.

[6] S. Boyd and S. S. Sastry. Necessary and sufficient conditions for parameter convergence in adaptive control. *Automatica*, 22:629–639, 1986.

[7] K. Doya and S. Yoshizawa. Adaptive neural oscillator using continuous time back-propagation. *Neural Networks*, 2:375–385, 1989.

[8] G. C. Goodwin, P. J. Ramadge, and P. E. Caines. Discrete-time multivariable adaptive control. *IEEE Transactions on Automatic Control*, 25:449–456, 1980.

[9] J. K. Hale. *Ordinary Differential Equations.* Wiley & Sons, New York, 1969.

[10] A. P. Morgan and K. S. Narendra. On the stability of nonautonomous differential equations $\dot{x} = [A + B(t)]x$, with skew symmetric matrix $B(t)$. *SIAM J. Control Optim.*, 15:163–176, 1977.

[11] K. Narendra and A. Annaswamy. Persistent excitation in adaptive systems. *Internat. J. Control*, 45:126–160, 1987.

[12] K. S. Narendra and A. M. Annaswamy. *Stable Adaptive Systems.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[13] B. A. Pearlmutter. Gradient calculations for dynamic recurrent neural networks: a survey. *IEEE Transactions on Neural Networks*, 6:1212–1228, 1995.

[14] R. Reinke. *Adaptive Regeln zum Lernen und Reproduzieren von periodischen Signalen mit dynamischen Netzwerken.* PhD thesis, Universität Kaiserslautern, 1994.

[15] S. Sastry and M. Bodson. *Adaptive Control: Stability, Convergence and Robustness.* Prentice Hall, London, 1979.

[16] L. Silverman and H. Meadows. Equivalent realizations of linear systems. *SIAM J. Appl. Math.*, 17:393–408, 1969.

[17] E. Sontag. *Mathematical Control Theory.* Springer Verlag, New York, 1990.

[18] M. G. Weiß. *Periodic signals in recurrent neural networks.* PhD thesis, Universität Kaiserslautern, ??

[19] J. Yuan and W. M. Wonham. Probing signals for model reference identification. *IEEE Transactions on Automatic Control*, 22:530–538, 1977.