Dariusz Jabłoński (Kraków)

THE CONJUGACY BETWEEN CASCADES GENERATED BY A WEAKLY NONLINEAR SYSTEM AND THE EULER METHOD OF A FLOW

Abstract. Sufficient conditions for the existence of a topological conjugacy between a cascade obtained from a weakly nonlinear flow by fixing the time step and a cascade obtained by the Euler method are analysed. The aim of this paper is to provide relations between constants in the Fečkan theorem. Given such relations an implementation of a weakly nonlinear neuron is possible.

1. Introduction. In recent years several papers have been devoted to the study of the qualitative properties of discrete time dynamical systems obtained via discretization methods. The basic question is whether the qualitative properties of a given continuous time dynamical system are preserved under discretization. Therefore many papers concern the problem of the existence of a topological conjugacy between a discretization of a flow and cascades generated by numerical methods [1, 3, 4]. In Fečkan's paper [3], a topological conjugacy between the discretization of the flow generated by a weakly perturbated linear differential equation and its discretization obtained by the Euler method is analysed. Fečkan's theorem ensures that there exists a topological conjugacy between the discretization of the flow induced by the differential equation $dx/dt = Ax + g(x)$ and the discrete dynamical system generated by the Euler method applied to this equation. The proof of the theorem is similar to the proof of the Grobman–Hartman theorem [5–7], known in the discrete and continuous cases. The discrete case of the Grobman–Hartman theorem yields the existence of a topological conjugacy between two small Lipschitz perturbations of a hyperbolic linear homeomorphism [5, 6]. Relations (6), (7) given below between constants in the Fečkan theorem constitute the first step towards the implementation of

a weakly nonlinear neuron because they are needed to introduce the proper form of the activation function of a weakly nonlinear neuron.

2. Preliminaries. Let $n \geq 2$ be a fixed positive integer. We define $\mathcal{L}(\mathbb{R}^n)$ to be the set of all linear endomorphisms of the space $\mathbb{R}^n$. Throughout this paper we consider $\mathcal{L}(\mathbb{R}^n)$ as a normed space, with the standard operator norm related to the Euclidean norm on $\mathbb{R}^n$. We say that an automorphism $T \in \mathcal{L}(\mathbb{R}^n)$ is *hyperbolic* if $\sigma(T) \cap S^1 = \emptyset$, where $S^1 \subset \mathbb{C}$ denotes the unit circle, and $\sigma(T)$ is the spectrum of the operator T. We define $\mathcal{C}_{\mathrm{B}}(\mathbb{R}^n)$ to be the set of all bounded continuous maps $\mathbb{R}^n \to \mathbb{R}^n$. Two continuous maps $\phi, \psi : \mathbb{R}^n \to \mathbb{R}^n$ are said to be *topologically conjugate* if there is a homeomorphism $\chi : \mathbb{R}^n \to \mathbb{R}^n$ such that $\psi = \chi^{-1} \circ \phi \circ \chi$. A *dynamical system* on $\mathbb{R}^n$ with continuous time is a family $\{\phi_t : t \in \mathbb{R}\}$ of homeomorphisms of $\mathbb{R}^n$ such that the map $(t, x) \mapsto \phi_t(x)$ is continuous, $\phi_0(x) = x$ and $\phi_{t+s}(x) = (\phi_t \circ \phi_s)(x)$ for all $x \in \mathbb{R}^n$ and all $t, s \in \mathbb{R}$. We define a *flow* to be a dynamical system on $\mathbb{R}^n$ with continuous time generated by a homogeneous differential equation [1]. If an $n \times n$ matrix A has no strictly imaginary eigenvalues, then the linear operator $A_h := \mathrm{id} + hA$ with a fixed $0 < h < \|A\|^{-1}$ is hyperbolic, which means that there exists an invariant splitting $\mathbb{R}^n = E^{\mathrm{s}} \oplus E^{\mathrm{u}}$ such that $\|A^{\mathrm{s}}\| \leq a < 1$, $\|(A^{\mathrm{u}})^{-1}\| \leq a < 1$, where $A^{\mathrm{s}} = A|E^{\mathrm{s}} : E^{\mathrm{s}} \to E^{\mathrm{s}}$ and $A^{\mathrm{u}} = A|E^{\mathrm{u}} : E^{\mathrm{u}} \to E^{\mathrm{u}}$. In the proof of the main result, the following lemma will be used (see [6, Lemma 4.3]).

LEMMA 2.1. *Let maps $\phi_1, \phi_2 \in \mathcal{C}_{\mathrm{B}}(\mathbb{R}^n)$ be Lipschitzian with constants smaller than $\varepsilon > 0$. Assume that the operator $A \in \mathcal{L}(\mathbb{R}^n)$ has no eigenvalue on the imaginary axis. Then the maps $A_h + \phi_1$ and $A_h + \phi_2$ are topologically conjugate provided*

$$\frac{\varepsilon \|A_h^{-1}\|}{1 - M} < 1,$$

where $M = \max\{\|(A^{\mathrm{u}})^{-1}\|, \|A^{\mathrm{s}}\|\}$.

3. The main result. Recall the Fečkan theorem (see [3]). Assume that Φ is the flow generated by the equation

$$(1) \qquad \frac{dx}{dt} = Ax + g(x),$$

where $A \in \mathcal{L}(\mathbb{R}^n)$ has no eigenvalues on the imaginary axis, $g \in \mathcal{C}_{\mathrm{B}}^1(\mathbb{R}^n)$ is such that $g(0) = 0$ and $\|\mathrm{D}g(x)\| \leq b$ for all $x \in \mathbb{R}^n$ and for a positive constant b. For the equation $dx/dt = Ax$ and a fixed $0 < h < \|A\|^{-1}$ we consider the discretization

$$(2) \qquad G(h, x) = A_h x, \qquad x \in \mathbb{R}^n,$$

generated by the Euler method. It is obvious that $G(h, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ is

a hyperbolic endomorphism for each $0 < h < \|A\|^{-1}$. We also consider the discretization

$$(3) \qquad H(h, x) = A_h x + h g(x), \qquad x \in \mathbb{R}^n,$$

generated by the Euler method for the equation (1). The Fečkan theorem can be expressed in the following way.

THEOREM 3.1. *For a sufficiently small $b > 0$ there is a number $h_0 > 0$ and*

$$(4) \qquad \Gamma : (0, h_0) \times \mathbb{R}^n \to \mathbb{R}^n$$

such that for all $h \in (0, h_0)$, $\Gamma(h, \cdot)$ is a homeomorphism satisfying

$$(5) \qquad \Phi(h, \cdot) \circ \Gamma(h, \cdot) = \Gamma(h, \cdot) \circ G(h, \cdot).$$

The aim of this paper is to give relations between the constants b and h in the Fečkan theorem:

THEOREM 3.2. *Let $0 < h < \|A\|^{-1}$ be fixed. Under the notation of the Fečkan theorem, if the following inequalities are true:*

$$(6) \qquad h \cdot b < (1 - M) \cdot \|A_h^{-1}\|^{-1},$$

$$(7) \qquad h \cdot b \left(\|A\| + b\right) < (1 - M) \cdot (\|A_h^{-1}\| \cdot \|e^{Ah}\|)^{-1},$$

where $M = \max\{\|(A^{\mathrm{u}})^{-1}\|, \|A^{\mathrm{s}}\|\}$, then the conclusion of the Fečkan theorem holds.

Proof. Due to the transitivity of the topological conjugacy relation, the proof can be divided into three steps.

STEP I. The following lemma has been proved in [6].

LEMMA 3.3. *Let $F(h, x) = e^{hA}x$, $x \in \mathbb{R}^n$, and let $\Phi(h, x)$ be the discretization of the flow generated by equation (1). Then the maps $\Phi(h, \cdot)$ and $F(h, \cdot)$ are topologically conjugate provided*

$$hb \cdot \|e^{Ah}\| \cdot (\|A\| + b) < (1 - M) \cdot \|A_h^{-1}\|^{-1}.$$

This yields the inequality (7).

STEP II.

LEMMA 3.4. *If $F(h, x) = e^{hA}x$ and $G(h, x) = A_h x$, then the maps $F(h, \cdot)$ and $G(h, \cdot)$ are topologically conjugate.*

Proof. We have

$$\#\{\nu \in \sigma(e^{hA}) \mid \nu > 1\} = \#\{\nu \in \sigma(A_h) \mid \nu > 1\},$$

where $\#$ denotes cardinality. The analogous equality holds for the eigenvalues smaller than 1. The assertion follows from [6]. ∎

STEP III.

LEMMA 3.5. *Let $g \in \mathcal{C}^1_{\mathrm{B}}(\mathbb{R}^n)$ be such that $\|\mathrm{D}g(x)\| < b$ for all $x \in \mathbb{R}^n$ and a constant $b > 0$. Then the maps $G(h, \cdot)$ and $H(h, \cdot)$ are topologically conjugate provided*

$$hb < (1 - M) \cdot \|A_h^{-1}\|^{-1}.$$

Proof. Recall that $G(h, x) = A_h x$ and $H(h, x) = A_h x + hg(x)$. Since $\|\mathrm{D}g(x)\| \leq b$ for all $x \in \mathbb{R}^n$, the map hg is Lipschitzian with constant $\varepsilon = hb$. So, according to Lemma 2.1, the map $G(h, \cdot)$ is topologically conjugate to $H(h, \cdot)$ provided

$$hb(1 - M)^{-1} < \|A_h^{-1}\|^{-1}. \quad \blacksquare$$

The condition (6) follows from Lemma 3.5. This completes the proof of Theorem 3.2.

4. Remarks. In general, it is not easy to compute eigenvalues of an operator. We shall give relations for the constants of the Fečkan theorem in the case where the matrix A_h is diagonal. Let k be a fixed positive integer such that $k < n$. We fix positive numbers $\lambda_1, \ldots, \lambda_k$ and negative numbers $\lambda_{k+1}, \ldots, \lambda_n$. Let

$$A = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k, \ldots, \lambda_n).$$

Then the matrix $A_h = \mathrm{id} + hA$ has the diagonal form

$$\begin{bmatrix} A^{\mathrm{u}} & 0 \\ 0 & A^{\mathrm{s}} \end{bmatrix},$$

where

$$A^{\mathrm{u}} = \mathrm{Diag}(1 + \lambda_1 h, \ldots, 1 + \lambda_k h),$$
$$A^{\mathrm{s}} = \mathrm{Diag}(1 + \lambda_{k+1} h, \ldots, 1 + \lambda_n h).$$

The following operator norms (see also Theorem 3.2) can be easily computed:

$$\|(A^{\mathrm{u}})^{-1}\| = \max\{(1 + \lambda_1 h)^{-1}, \ldots, (1 + \lambda_k h)^{-1}\},$$
$$\|A_h^{-1}\| = \min\{1 + \lambda_{k+1} h, \ldots, 1 + \lambda_n h\},$$
$$\|A^{\mathrm{s}}\| = \max\{1 + \lambda_{k+1} h, \ldots, 1 + \lambda_n h\},$$
$$\|A_h\| = \max\{1 + \lambda_1 h, \ldots, 1 + \lambda_k h\},$$
$$\|A\| = \max\{|\lambda_1|, \ldots, |\lambda_k|, \ldots, |\lambda_n|\},$$
$$\|e^{hA}\| = \max\{e^{\lambda_1 h}, \ldots, e^{\lambda_k h}\}.$$

To check whether the topological conjugacy between the flow induced by a weakly perturbed linear differential equation and its discretization obtained via the Euler method exists (Theorem 3.2) it is enough to estimate the norms of the matrices. In numerical applications computing the eigenvalues of the matrix A allows us to calculate the norms of the operators

$A^{\mathrm{s}}, (A^{\mathrm{u}})^{-1}, A_h, A_h^{-1}, e^{hA}$. Next we have to choose a map g with Lipschitz constant b such that the inequalities of Theorem 3.2 hold.

5. Applications to artificial neural networks. The relations between the constants in the Fečkan theorem are applicable to the mathematical basis of artificial neural networks. Let us recall the basic definitions (see [2]). An artificial neuron is a unit having several weighted inputs and one output (Figure 1). Thus we can say that a neuron is a function of two vector variables. A *neuron* with l inputs transforming a set $X \subset \mathbb{R}$ of input signals (an l-neuron on X) is a function

$$F : \mathbb{R}^l \times X^l \ni (w, x) \mapsto F(w, x) = f(\langle w, x \rangle) \in \mathbb{R},$$

where $\langle \cdot, \cdot \rangle$ is a real scalar product and $f : \mathbb{R} \to \mathbb{R}$ is called the *activation function* of the neuron. If f is linear, then the neuron is said to be *linear*.
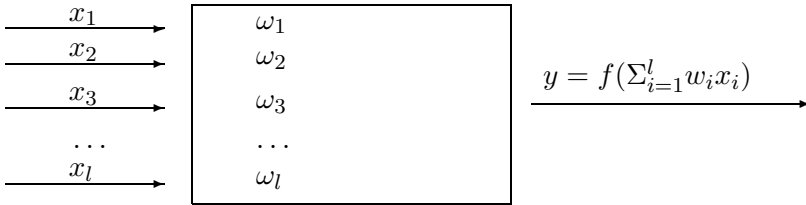

$$
\begin{array}{ccc}
x_1 & \omega_1 & \\
x_2 & \omega_2 & \\
x_3 & \omega_3 & y = f(\Sigma_{i=1}^l w_i x_i) \\
\dots & \dots & \\
x_l & \omega_l &
\end{array}
$$

Fig. 1. A neuron model

Neuron weights are set during a training process. Thus a *trained neuron* can be defined as a mapping

$$(8) \qquad F^\star = F(w, \cdot) : \mathbb{R}^l \to \mathbb{R}.$$

An *artificial neural network* is a system of neurons which are connected in such a way that an output signal of one neuron is transmitted to the input of another neuron. The most often considered neural networks are those with a layer structure of connections. Multilayer ANNs (Artificial Neural Networks) should act correctly. This means that in response to a given input signal the output signal should be a desired value. If some of these desired output values are known, a training sequence, used in a training process, can be defined. A finite sequence of pairs

$$((\vec{x}^{(1)}, z^{(1)}), \ldots, (\vec{x}^{(N)}, z^{(N)})),$$

where $\vec{x}^{(i)}$ are input signals of an l-ANN on X and $z^{(i)}$ is a required output signal, is called a *training sequence* of the multilayer l-ANN on X. Adjustment of weights causing the differences between real and desired outputs in response to a given input signal to be as small as possible is called a *training process* of an ANN. The problem of weight adjustment can be reduced to finding the minimum of a certain function, called the *criterial function*. We

will consider one-layer networks. Then the learning process of a neuron is independent of learning others. In such a case the most often used criterial function is given by

$$(9) \qquad E(w_1, \ldots, w_M) = \sum_{n=1}^{N} [y(w_1, \ldots, w_M)^{(n)} - z^{(n)}]^2,$$

where

$$(10) \qquad y(w_1, \ldots, w_M)^{(n)} = f\Big(\sum_{m=1}^{M} x_m^{(n)} w_m \Big).$$

Here M is the number of neuron inputs and N is the number of elements in a training sequence. The *descent gradient learning process* of a neuron is given by the iterative rule

$$(11) \qquad w_t^{(p+1)} = w_t^{(p)} - \eta \cdot \frac{\partial \mathrm{E}(\mathbf{w}^{(p)})}{\partial w_t},$$

where the index p is the number of iterations and $\mathbf{w} = \{w_t\}_{t \in T}$ is the vector of all weights in the ANN. The formula (11) is the Euler method for the gradient differential equation

$$(12) \qquad \frac{dw_t}{ds} = -\mathrm{grad}\, \mathrm{E}(\mathbf{w}).$$

Thus the output deviation function plays the role of the potential E in the gradient equation (12). Since, for sufficiently small η, (11) generates a discrete dynamical system whereas (12) generates a continuous dynamical system, the dynamical system theory can be used in the training process analysis. We can see ([2]) that the gradient learning process of a weakly perturbed linear neuron is modelled by a weakly perturbed linear differential equation. A special type of an artificial neuron, called a *weakly nonlinear neuron*, was introduced by Bielecki [2]. A neuron is said to be *weakly nonlinear* if its activation function is of the form

$$(13) \qquad f : \mathbb{R} \ni \beta \mapsto f(\beta) = \beta + u(\beta),$$

where $u \in \mathcal{C}_{\mathrm{B}}^2(\mathbb{R})$ is such that $|u'(\beta)| < c_1$, $|u''(\beta)| < c_2$, $|\beta \cdot u'(\beta)| < c_3$, $|\beta \cdot u''(\beta)| < c_4$ for each $\beta \in \mathbb{R}$, where the constants c_1, c_2, c_3, c_4 are sufficiently small as well. It has been shown that if the gradient equation (12) is used to describe the training process of a weakly nonlinear neuron, i.e. the function f in (10) is of the type (13), then the equation (12) satisfies the assumptions of the Fečkan theorem. Since the constants c_2, c_3, c_4 depend on the constants b and h in the Fečkan theorem (see Theorem 3.1), the results obtained are necessary to introduce actual forms of the mapping u (see formula (13)) and to start implementation experiments with weakly nonlinear neurons.

References

[1]   N. Aoki and K. Hiraide, *Topological Theory of Dynamical Systems*, North-Holland, Tokyo, 1994.

[2]   A. Bielecki, *Dynamical properties of learning process of weakly linear and nonlinear neurons*, Nonlinear Anal. Ser. B 2 (2001), 249–258.

[3]   M. Fečkan, *The relation between a flow and its discretization*, Math. Slovaca 42 (1992), 123–127.

[4]   B. Garay, *On structural stability of ordinary differential equations with respect to discretization methods*, J. Numer. Math. 4 (1996), 449–479.

[5]   M. C. Irwin, *Smooth Dynamical Systems*, Academic Press, New York, 1980.

[6]   J. Palis and W. de Melo, *Geometric Theory of Dynamical Systems*, Springer, New York, 1982.

[7]   M. Shub, *Global Stability of Dynamical Systems*, Springer, New York, 1986.

Institute of Mathematics
Cracow University of Technology
Warszawska 24
31-155 Kraków, Poland
E-mail: djablons@riad.usk.pk.edu.pl