

Vector Fields and Neural Networks

R. Vilela Mendes

Theoretical Division, CERN, CH-1211, Geneva 23, Switzerland
and

*Centro de Física da Matéria Condensada, Av. Gama Pinto,
2-1699 Lisboa Codex, Portugal*

J. Taborda Duarte

*Laboratório Nacional de Engenharia e Tecnologia Industrial,
Az. dos Lameiros, Estr. do Paço do Lumiar, 1600 Lisboa, Portugal*

Abstract. We consider neural network models described by systems of (continuous time) differential equations. The dynamical nature of each model is identified, symmetric networks being related to gradient vector fields and asymmetric networks decomposed into their gradient and Hamiltonian components. From this identification follows, in particular, a simple characterization of structural stability for symmetric networks and a limit cycle analysis of asymmetric networks as generators of coherent temporal patterns.

1. Introduction

Although the signals sent by neurons along their axons are sharp spikes, the relevant information is probably not contained in the spikes themselves but in the firing rate, which depends on the magnitude of the membrane potential. This in turn results from the integrated effect of the firing rates of the presynaptic neurons. In this sense a continuous time model with continuous state variables might be closer to an actual neural system than a binary discrete time model, at least for certain areas of the brain. It should be pointed out, however, that the recent neurophysiological observations of extremely low firing rates [1] sheds some doubt on the usefulness of the firing rate as the relevant neural variable.

Because of the convenience of electronic implementation, there is a bias toward binary systems in neural network simulations. However, by smoothing the energy surfaces and eliminating shallow local minima, the use of continuous degrees of freedom improves in general the speed of convergence to near-optimal solutions in associative memory, optimization (annealing) problems, and supervised learning. As pointed out by Hopfield and Tank [2], analog parallel dynamics, as opposed to digital sequential dynamics, allows

all the elements of the network to cooperate simultaneously in the computed decision. Also, during the computation, intermediate values of the variables are explored, which in a logical sense would correspond to the use of propositions that are neither wholly true nor wholly false. That is, there is a logical broadening of the nature of the computing process. In this sense the brain mode of operation, in which neurons react to the average of a sequence of spikes, may be close to ideal because it benefits from the features of continuous parallel computation and, at the same time, the spikes provide the amount of fluctuation, which allows for evasion from shallow local minima.

A number of authors [2, 3] were thus led to use continuous variables to describe neural networks and bring the techniques of non-linear differential equations to the formulation of their dynamical problems. In this paper we somehow pursue this trend in the sense that, starting from a general system that contains many of the proposed models as particular cases, we characterize a network's behavior by identifying its nature as a dynamical system.

It is a simple matter to check the relation between neural networks with symmetric synaptic strengths and gradient dynamical systems. We extend this analysis to non-symmetric networks exhibiting a decomposition into gradient and Hamiltonian components.

The identification of the dynamical classes, where the neural network models fit, provides a methodology for a systematic application of the results of the general theory of dynamical systems to this field. In particular, results concerning structural stability as well as perturbation and averaging techniques seem to be promising.

Although the emphasis is, as stated, on analog neural networks, most of the results and dynamical analysis also apply to the equations for the mean local magnetization [4] of discrete neurons.

2. Gradient fields and Lyapunov functions

Cohen and Grossberg realized in 1983 [5] that many models of content addressable memories (CAM), as well as other systems, can be written in the form

$$\frac{dx_i}{dt} = a_i(x^i) \left\{ b_i(x^i) - \sum_{j=1}^n c_{ij} d_j(x^j) \right\} \quad (2.1a)$$

and if

$$c_{ij} = c_{ji} \quad (2.1b)$$

and

$$a_i(x^i) d'_i(x^i) \geq 0 \quad (2.1c)$$

there is a Lyapunov function

$$V = - \sum_{i=1}^n \int^{x^i} b_i(\xi^i) d'_i(\xi^i) d\xi^i + \frac{1}{2} \sum_{j,k=1}^n c_{jk} d_j(x^j) d_k(x^k) \quad (2.2)$$

that is, $dV/dt \leq 0$ along the orbits.

The dynamical evolution toward the stationary points of the Lyapunov function defines the memory states learned by or built into the network, as well as their basins of attraction. This Lyapunov function has since been used and rediscovered in the context of several models.

The dynamical behavior of CAM models of this type is closely related to the behavior of a gradient field. Assuming differentiability of the functions involved in (2.1a), one easily finds that the dynamical system (2.1) is differentially equivalent to a gradient system. That is, there is a differentiable change of coordinates $y^i = y^i(x)$ such that the system reduces to

$$\frac{dy^i}{dt} = -\frac{\partial V}{\partial y^i} \tag{2.3}$$

In fact, from (2.1) and (2.2) one obtains

$$\frac{dx^i}{dt} = -\frac{a_i(x^i)}{d'_i(x^i)} \frac{\partial V}{\partial x^i} \tag{2.4}$$

and the change of coordinates leading to (2.3) is

$$y^i(x^i) = \int^{x^i} \left(\frac{d'_i(\xi^i)}{a_i(\xi^i)} \right)^{1/2} d\xi^i \tag{2.5}$$

where in (2.1c) we now require strict positivity.

Alternatively, from (2.4) we could state that \dot{x}^i itself is a gradient vector field in the metric

$$g^{ij}(x) = \frac{a_i(x^i)}{d'_i(x^i)} \delta^{ij} \tag{2.6}$$

The Lyapunov function of the Cohen-Grossberg model is of a special type because not all vector fields with Lyapunov functions are necessarily equivalent to gradient fields. On the other hand, it is clear that other models exist that are differentially equivalent to gradient fields and are not of the Cohen-Grossberg type. In general, given an arbitrary function $V(x)$ of n coordinates and a set of n functions $y^i = y^i(x)$, $i = 1, \dots, n$ with non-vanishing Jacobian $(\partial y / \partial x)$, there is a dynamical system

$$\frac{dx^i}{dt} = -\sum_k \left\{ \left(\frac{\partial y}{\partial x} \right)^{-1} \left(\frac{\partial y}{\partial x} \right)^{-1T} \right\}_{ik} \frac{\partial V}{\partial x^k} \tag{2.7}$$

having $V(x)$ as its Lyapunov function.

The relation between general CAM models and gradient fields is not of purely academic interest. The differential equivalence of Cohen-Grossberg symmetric models to gradient vector fields means that a complete characterization of these models may be obtained as a straightforward application of the known properties of gradient fields [6]. For example, for a dynamical

system to be an adequate model for a neural system it should be robust, in the sense that small changes in the parameters should result only in small changes in the qualitative behavior of the solutions. This corresponds to the notion of structural stability, which requires that our system possess a C^1 -neighborhood of systems topologically equivalent to the original one by homeomorphisms close to the identity.

A necessary condition for structural stability of gradient vector fields is the non-degeneracy of the critical points of V , namely

$$\det \left\| \frac{\partial^2 V}{\partial x^i \partial x^j} \right\| \neq 0 \text{ at the points where } \frac{\partial V}{\partial x^i} = 0.$$

In a gradient flow, all orbits approach the critical points as $t \rightarrow \infty$. If the critical points are non-degenerate, then the gradient flow satisfies the conditions defining a Morse-Smale field, except perhaps the transversality conditions for stable and unstable manifolds of the critical points. However, because Morse-Smale fields are open and dense in the set of gradient vector fields, any gradient flow with non-degenerate critical points may always be C^1 -approximated by a (structurally stable) Morse-Smale gradient field.

3. Hamiltonian components and coherent temporal patterns

Synaptic connections in biological neural networks are seldom symmetric. Quite often only one of c_{ij} or c_{ji} is nonzero. It is therefore important to identify the nature of the dynamical systems that are implemented by asymmetric neural networks. Because of its generality we use once again the Cohen-Grossberg form (2.1), but now the c_{ij} 's are not required to be symmetric. We consider the dynamical system to have an even number n of variables. Otherwise it is trivially embedded into a system of dimension $n + 1$. Define

$$\begin{aligned} c_{ij} &= c_{ij}^{(S)} + c_{ij}^{(A)} \\ c_{ij}^{(S)} &= \frac{1}{2}(c_{ij} + c_{ji}) & c_{ij}^{(A)} &= \frac{1}{2}(c_{ij} - c_{ji}) \end{aligned} \quad (3.1)$$

$$V^{(S)} = - \sum_{i=1}^n \int^{x^i} b_i(\xi^i) d'_i(\xi^i) d\xi^i + \frac{1}{2} \sum_{j,k=1}^n c_{jk}^{(S)} d_j(x^j) d_k(x^k) \quad (3.2)$$

and

$$H = \sum_{i=1}^n \int^{x^i} \frac{d_i(\xi^i)}{a_i(\xi^i)} d\xi^i \quad (3.3)$$

Then we have the following.

Theorem 1. *If $a_i(x^i)/d'_i(x^i) > 0 \forall x, i$ and the matrix $c_{ij}^{(A)}$ has an inverse, then the vector field \dot{x}^i in equation (2.1a) decomposes into gradient and Hamiltonian components, $\dot{x}^i = \dot{x}^{i(G)} + \dot{x}^{i(H)}$, where*

$$\dot{x}^{i(G)} = - \frac{a_i(x^i)}{d'_i(x^i)} \frac{\partial V^{(S)}}{\partial x^i} = - \sum_j g^{ij}(x) \frac{\partial V^{(S)}}{\partial x^j} \quad (3.4a)$$

$$\dot{x}^{i(H)} = - \sum_j a_i(x^i) c_{ij}^{(A)}(x) a_j(x^j) \frac{\partial H}{\partial x^j} = \sum_j I^{ij}(x) \frac{\partial H}{\partial x^j} \quad (3.4b)$$

and

$$g^{ij}(x) = \frac{a_i(x^i)}{d'_i(x^i)} \delta^{ij} \quad \text{and} \quad \omega_{ij} = -a_i(x^i)^{-1} \left(c_{ij}^{(A)-1} \right)_{ij} a_j(x^j)^{-1},$$

where $\omega_{ij} I^{jk} = \delta_i^k$ are the components of the Riemannian metric and the symplectic form.

Proof. The decomposition follows by direct calculation from (2.1a), (3.2), and (3.3). The conditions on $a_i(x^i)$, $d'_i(x^i)$, and $c_{ij}^{(A)}$ insure that g is a well-defined metric and ω is non-degenerate. Indeed let v^i be a vector such that $\sum_i v^i \omega_{ij} = 0$. Then

$$0 = \sum_{i,j} v^i \omega_{ij} a_j(x^j) c_{jk}^{(A)} = - \frac{v^k}{a_k(x^k)}$$

would imply $v^k = 0 \forall k$. That ω is a closed form follows from the fact that ω_{ij} depends only on x^i and x^j . ■

The identification in the Cohen-Grossberg model of just one gradient and one Hamiltonian component with explicitly known potential and Hamiltonian functions is a considerable simplification over a generic dynamical system. We recall that in the general case, although such a decomposition is possible locally [7], explicit functions are not easy to obtain unless one allows for one gradient and $n - 1$ Hamiltonian components. Notice that the decomposition of the vector field does not decouple the dynamical evolution of the components. In fact, as seen below, it is the interplay of the dissipative (gradient) and the Hamiltonian components that leads to the limit cycle behavior.

In light of the non-negligible fact that actual neural connections are asymmetric, the main motivation for studying non-symmetric networks is the potential for understanding the emergence of coherent temporal behavior. Whereas in CAM symmetric networks the memories that are recalled are the minima of the Lyapunov function, in the asymmetric networks the memories might be stable oscillations or non-periodic temporal sequences. In both cases these patterns might be used to process temporal information.

It is clear that the behavior of an asymmetric network will depend on the relative size of the gradient and the Hamiltonian components. If for example the Hamiltonian component is very small, then the dynamics will be determined primarily by the gradient part, the trajectories flowing to the neighborhood of the minima of V . We will be mostly concerned, however, with the cases where H is sufficiently large to determine stable persistent motions in some regions of phase space.

When in a non-conservative system persistent motions are found in some subspace of the whole phase space, it means that dissipation and regeneration

effects compensate each other in such a way that a kind of “local energy” or constant of motion is preserved along some orbits. A well known case [8] is the creation of limit cycles in the plane from closed orbits of perturbed Hamiltonian systems. The necessary condition for the existence of the cycle is the vanishing of the variation of the monodromy

$$I(c) = \oint (B dx - A dy) = 0 \quad (3.5)$$

with

$$\dot{x} = \frac{\partial H}{\partial y} + \epsilon A(x, y, \epsilon) \quad \dot{y} = -\frac{\partial H}{\partial x} + \epsilon B(x, y, \epsilon)$$

In reference [9] this idea has been generalized by introducing the notions of “constant of motion” and “arc of vector fields with constants of motion.” We recall these definitions:

Let (M, X) be a differentiable dynamical system, where M is a manifold and X a smooth vector field. A “constant of motion” of (M, X) is any differentiable function $\phi : M \rightarrow \mathbf{R}$ such that for *some* solution (orbit) γ of X we have $\phi \circ \gamma = \text{constant}$. Notice the qualification *some* rather than *all*, which would have been the case had ϕ been a first integral.

A family of vector fields $\epsilon \rightarrow X_\epsilon$, $\epsilon \in [-a, a]$ is called an “arc of vector fields with constants of motion” if

1. Each X has a constant of motion ϕ_ϵ over a periodic solution γ_ϵ .
2. The constant of motion ϕ_o of X_o is a first integral in a neighborhood of γ_o .
3. The maps $\epsilon \rightarrow X_\epsilon$, $\epsilon \rightarrow \gamma_\epsilon$, and $\epsilon \rightarrow \phi_\epsilon$ are C^1 -differentiable.

The main result of reference [9] is a generalization of equation (3.5), stating a necessary condition for the existence of the arcs, namely for local persistent dynamics with a constant of motion. This is

$$\int_o^{T_o} i \left(\frac{dX_\epsilon}{d\epsilon} \Big|_{\epsilon=o} \right) (d\phi_o) \gamma_o(t) dt = 0 \quad (3.6)$$

where T_o is the period of the orbit of X_o .

Because of condition 2, the point $\epsilon = 0$ in the arc where the derivative is computed is a point where the corresponding vector field has a first integral. Therefore, although the result (3.6) is very general, in practice it is useful mostly when X_o is a Hamiltonian field because then at least one first integral, the energy, is known.

Applying (3.6) to the system (2.1a), and defining the arc as

$$X_\epsilon = \epsilon \dot{x}^{(G)} + \dot{x}^{(H)} \quad (3.7)$$

with $\dot{x}^{(G)}$ and $\dot{x}^{(H)}$ as in Theorem 1 and the constant of motion ϕ_o being H itself, one obtains

$$\begin{aligned} \int_o^T \sum_i \left(\frac{\partial H}{\partial x^i} a_i(x^i) \frac{\partial V^{(S)}}{\partial x^i} \right) dt \\ = 0 = \int_o^T \sum_i d_i(x^i) \left\{ b_i(x^i) - \sum_{i,j} c_{ij}^{(S)} d_j(x^j) \right\} dt \end{aligned} \quad (3.8)$$

the integration being taken along a closed orbit of H .

Example 1. As a first example consider a network with two groups of neurons, which we denote by the coordinates $\{x\}$ and $\{y\}$, connected in such a way that inside both groups the synaptic strengths are symmetric, whereas between groups they are antisymmetric. That is, $\{x\}$ is excitatory to $\{y\}$, and $\{y\}$ is inhibitory to $\{x\}$.

Assume the self-regulation functions b_i of the neurons to be

$$b(x) = \beta_1 x - \beta_3 x^3 \quad \beta_1, \beta_3 > 0 \quad (3.9a)$$

$$b(y) = \beta_4 y - \beta_6 y^3 \quad \beta_4, \beta_6 > 0 \quad (3.9b)$$

where β_1, β_4 and β_3, β_6 are auto-excitation and self-saturation coefficients, respectively.

For simplicity, let $a_i(x) = a_i(y) = 1$, $d(x) = x$, $d(y) = y$, and the synaptic strengths between the $\{x\}$ and $\{y\}$ groups have constant amplitude ($c^{(yx)} = -c^{(xy)} = c$). The symmetric strengths $c^{(xx)}$ and $c^{(yy)}$ inside the groups $\{x\}$ and $\{y\}$ are left to be adjusted by some learning process. Then the potential function $V^{(S)}$ of the gradient component is

$$\begin{aligned} V^{(S)} = \sum_i -\frac{\beta_1}{2} (x^i)^2 + \frac{\beta_3}{4} (x^i)^4 + \frac{1}{2} \sum_{j,k} c_{jk}^{(xx)} x^j x^k \\ + \sum_i -\frac{\beta_4}{2} (y^i)^2 + \frac{\beta_6}{4} (y^i)^4 + \frac{1}{2} \sum_{j,k} c_{jk}^{(yy)} y^j y^k \end{aligned} \quad (3.10a)$$

and the Hamiltonian is

$$H = \frac{1}{2} \sum (x^i)^2 + \frac{1}{2} \sum_i (y^i)^2 \quad (3.10b)$$

Notice however that, whereas the metric is trivial ($g^{ij} = \delta^{ij}$), the inverse of the symplectic form connects each x^i with all y 's, and conversely.

The equation of motion corresponding to the Hamiltonian H , namely the dynamics in the limit $\epsilon = 0$ in (3.7), is

$$\dot{x}_o^i = -c \sum_j y_o^j \quad (3.11a)$$

$$\dot{y}_o^i = c \sum_k x_o^k \quad (3.11b)$$

From (3.11) one concludes that the Hamiltonian orbits are parametrized by

$$x_o^i = A \cos \omega t \quad (3.12a)$$

$$y_o^i = B \sin \omega t \quad (3.12b)$$

with $\omega^2 = c^2 N_x N_y$ and $B = A \sqrt{N_x/N_y}$, with N_x and N_y being the number of neurons in the groups.

Using equation (3.8) one then obtains

$$A^2 = \frac{\beta_1 + \beta_4 - (1/N_x) \sum_{ij} c_{ij}^{(xx)} - (1/N_y) \sum_{ij} c_{ij}^{(yy)}}{\frac{3}{4}(\beta_3 + (N_x/N_y)\beta_6)} \quad (3.13)$$

as the approximate amplitude of the limit cycle. One sees that the frequency of the persistent oscillation is determined by asymmetric synaptic strengths, whereas its amplitude is regulated by the symmetric strengths inside the $\{x\}$ and $\{y\}$ groups.

We want to point out that equation (3.8) gives only an approximate estimate of the limit cycle, whenever it exists. However, as shown in reference [9] by numerical simulation, in most cases equation (3.8) provides a good qualitative estimate of the persistent motions even for large values of the deformation parameter.

Example 2. For the second example we take neurons with self-regulation functions b_i of the form

$$b(x) = -\beta_1 x + \beta_2 x^3 - \beta_3 x^5 \quad \beta_1, \beta_2, \beta_3 > 0 \quad (3.14a)$$

$$b(y) = -\beta_4 y + \beta_5 y^3 - \beta_6 y^5 \quad \beta_4, \beta_5, \beta_6 > 0 \quad (3.14b)$$

and consider several groups of $\{x\}$ and $\{y\}$ neurons, as in the preceding example, with the asymmetric connections and the number of neurons chosen in such a way that each group $\Gamma_k = [\{x\}_k, \{y\}_k]$ has a different Hamiltonian frequency $\omega_k = c_k (N_x^k N_y^k)^{1/2}$.

The self-regulation functions (3.14) lead to a potential

$$\begin{aligned} V^{(S)} = & \sum_i \frac{\beta_1}{2} (x^i)^2 - \frac{\beta_2}{4} (x^i)^4 + \frac{\beta_3}{6} (x^i)^6 + \frac{1}{2} \sum_{j,k} c_{jk}^{(xx)} x^j x^k \\ & + \sum_i \frac{\beta_4}{2} (y^i)^2 - \frac{\beta_5}{4} (y^i)^4 + \frac{\beta_6}{6} (y^i)^6 + \frac{1}{2} \sum_{j,k} c_{jk}^{(yy)} y^j y^k \end{aligned}$$

We now have two stable modes of operation in each neuron group Γ_k . One is the quiescent state with all $x^i = y^i = 0$, and the other is a limit cycle of approximate frequency ω_k and amplitude

$$A_k^2 = \frac{\beta_2 + \beta_5 (N_x/N_y) + \Delta}{2(\beta_3 + (N_x/N_y)^2 \beta_6)}$$

with

$$\Delta = \left[\left(\beta_2 + \beta_5 \left(\frac{N_x}{N_y} \right) \right)^2 - 4 \left(\beta_3 + \left(\frac{N_x}{N_y} \right)^2 \beta_6 \right) \times \left(\beta_1 + \beta_4 + \frac{1}{N_x^k} \sum_{ij} c_{ij}^{(xx)} - \frac{1}{N_y^k} \sum_{ij} c_{ij}^{(yy)} \right) \right]^{1/2}$$

In particular, oscillations of small amplitude will die out and the network will tend to the quiescent state. This network may be a model for a system that, subject to an oscillating stimulus of sufficient intensity and an appropriate frequency, will excite one of the neuron groups if there is one with a frequency that closely matches the stimulus.

Using a similar approach it also possible to construct models of temporal association whereby a network, when triggered, excites sequentially a set of neurons. It is enough to construct a network with both a quiescent state and a stable Hamiltonian orbit that passes through the required states. Then, if triggered to a state belonging to the stable orbit, the network will unfold the required temporal pattern.

4. Final remarks

The decomposition into components with well-studied dynamical properties provides a strategy for further systematic work in neural network models. For example,

1. The bifurcation theory of gradient dynamical systems is certainly related to the mechanisms of creation of new fixed points in the learning process of associative memories.
2. Coupling a small Hamiltonian perturbation to a primarily dissipative system with many degrees of freedom, one obtains local fluctuations of the variables without losing the global coherence, in the sense that the motion remains near a global fixed point of the unperturbed dissipative system. This may be relevant to the construction of models where the mean active time of individual neurons is less than the memory coherence time.

In this paper we were concerned with the fast dynamics of the neural model. If, in addition, we interpret learning as a slow dynamical process coupled to the fast dynamics of the nodes, we conclude that the global system may have non-trivial behavior that arises from the coupling. Consider for example a learning dynamics of the Hebbian type, for example

$$\frac{dc_{ij}}{dt} = -\lambda_{ij}c_{ij} + K_{ij}w_i(x^i)z_j(x^j) \tag{4.1}$$

A passive decay term with coefficient λ_{ij} was considered and we have allowed for different functions w and z in the post- and pre-synaptic neurons. Other more complicated learning processes have been considered [3], allowing for

example the decay to be gated by the neuron variables. However, for the purposes of this remark, this simple form will suffice.

Typically in the learning process a set of neurons (conventionally denoted inputs and outputs) are fixed and the rest of the network allowed to relax to values $x^i(\infty)$. Then the neuron variables are fixed at $x^i = x^i(\infty)$ and the synaptic strengths updated according to the relaxation of (4.1). However, when the c_{ij} 's change, the variables x^i that are not fixed from outside will start to relax to new values. This will certainly occur if, as expected, the relaxation times of the x^i 's are smaller than those of the c_{ij} 's. What we have then is a coupled dynamical system with equations (2.1) and (4.1), plus constraints. (The constraints are the functional relations resulting from the externally fixed variables.) Even when the c_{ij} 's are symmetric, the coupled system is no longer differentially equivalent to a gradient system. It will have Hamiltonian components and oscillations become possible in the dynamics of the learning process.

References

- [1] M. Abeles, E. Vaadia, and H. Bergman, *Network*, **1** (1990) 13–25.
- [2] J. J. Hopfield and D. W. Tank, *Biological Cybernetics*, **52** (1985) 141; see also “Collective Computation with Continuous Variables,” pages 155–170 in *Disordered Systems and Biological Organization*, edited by E. Bienenstock et al. (Berlin, Springer, 1986).
- [3] S. Grossberg, *Neural Networks*, **1** (1988) 17–61, and references therein.
- [4] D. J. Amit, *Modeling Brain Function* (Cambridge, Cambridge University Press, 1989).
- [5] M. A. Cohen and S. Grossberg, *IEEE Transactions on Systems, Man and Cybernetics*, **13** (1983) 815–826.
- [6] S. Smale, *Annals of Mathematics*, **74** (1961) 199–206. J. Palis and S. Smale, *Proceedings of the Symposium on Pure Mathematics*, vol. XIV (Providence, American Mathematical Society, 1970).
- [7] R. Vilela Mendes and J. T. Duarte, *Journal of Mathematical Physics*, **22** (1981) 1420–1422.
- [8] See, for example, D. V. Anosov and V. I. Arnold, “Dynamical Systems,” in volume I of *Encyclopaedia of Mathematical Sciences* (Berlin, Springer, 1988).
- [9] J. T. Duarte and R. Vilela Mendes, *Journal of Mathematical Physics*, **24** (1983) 1772–1778; *Letters on Mathematical Physics*, **6** (1982) 249–252; for a review and generalizations of these results see also R. Vilela Mendes, “Deformation Stability of Periodic and Quasi-periodic Motion in Dissipative Systems,” in *Deformation Theory of Algebras and Structures and Applications*, edited by M. M. Hazewinkel and M. Gerstenhaber (Dordrecht, Kluwer Academic Publishers, 1988).