C. R. Acad. Sci. Paris, Ser. I 335 (2002) 639–642

Physique mathématique/*Mathematical Physics*

Complexité dynamique des réseaux de Hopfield

Serge Vakulenko

Institute for Mechanical Engineering Problems, Bolshoy pr. V.O. 61, St. Petersbourg, 199178 Russia

Reçu le 26 juillet 2001 ; accepté après révision le 3 juin 2002

Note présentée par Pierre-Louis Lions.

Résumé On considère les réseaux de neurones de Hopfield. On montre que ce système peut engendrer toute dynamique inertielle structurellement stable, avec mémoire bornée. ***Pour citer cet article : S. Vakulenko, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 639–642.***

Dynamical complexity of the Hopfield networks

Abstract One considers the Hopfield networks. It is shown that this system can generate any structurally stable inertial dynamics, with a bounded memory. ***To cite this article: S. Vakulenko, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 639–642.***

1. Introduction

L'objet de cette Note est le système de Hopfield : un modèle fondamental de la théorie des réseaux de neurones. Nous allons démontrer que ces systèmes possèdent une grande complexité dynamique : quand nous changeons leurs paramètres, ils peuvent engendrer une large classe de dynamiques inertielles avec retard. En particulier, ils engendrent tous les attracteurs structurellement stables et tous les ensembles hyperboliques (à un homéomorphisme près).

Remarquons que cette propriété ne peut pas être vérifiée par ordinateur, même très puissant. C'est un résultat purement analytique et, de plus, pour ces systèmes, nous donnons un algorithme constructif du contrôle de l'attracteur global.

Cette Note développe des résultats précédents [9,8], où l'existence d'une dynamique inertielle compliquée (sans retard) a été démontrée. Contrairement à [9,8], les résultats de cette Note sont valables quand les communications entre des neurones satisfont à des restrictions importantes qui sont biologiquement fondamentales, en particulier, tous les neurones sont identiques et ils n'agissent pas sur eux-mêmes. Donc, même sous de telles restrictions, la dynamique neuronale peut être pratiquement arbitraire.

2. Position du problème

Le modèle de Hopfield est le système dynamique suivant à temps discret :

$$x_i(t+1) = \sigma\left(\mathbf{K}_i x(t) + \theta\right), \tag{1}$$

où $\mathbf{x} = (x_1(t), \ldots, x_N(t)) \in \mathbf{R}^N$ est un vecteur d'état des neurones, N est le nombre de neurones. Notons $\mathbf{K}_i x = \sum_{j=1}^{N} K_{ij} x_j$ où $\mathbf{K}$ est une matrice définissant les communications entre les neurones, θ un paramètre

Adresse e-mail : vakul@mech.ipme.ru (S. Vakulenko).

S1631-073X(02)02524-4/FLA

(« seuil ») et $\sigma(z)$ une fonction « sigmoïdale » vérifiant les conditions suivantes :

$$\sigma'(z) \in S(\mathbf{R}), \quad \sigma'(z) > 0, \quad \sigma'(0) = 1, \tag{2}$$

où $S(\mathbf{R})$ désigne la classe de fonctions de L. Schwartz des fonctions indéfiniment dérivables et décroissant, pour $|z| \to \infty$, plus rapidement que tout puissance de $|z|^{-1}$ ainsi que chacune de leurs dérivées (un exemple typique est : $\sigma(z) = \tanh(z)$). De plus, nous supposons que

$$\mathbf{K} = \lambda\mathbf{M}, \quad M_{ij} = 1, -1 \text{ ou } 0, \quad M_{ii} = 0, \quad \lambda > 0. \tag{3}$$

Fixons une fonction σ. Considérons les nombres N, λ, la matrice $\mathbf{M}$ et θ comme l'ensemble des paramètres $\mathcal{P}$ de notre système, $\mathcal{P} = \{N, \lambda, \theta, \mathbf{M}\}$.

Introduisons la classe suivante des systèmes dynamiques à temps discret ($t \in \mathbf{N}$) et à mémoire bornée :

$$q(t) = F\big(q(t-2), q(t-4), \ldots, q(t-2\mu)\big), \tag{4}$$

où $q \in \mathbf{R}^n$, $\mu \in \mathbf{N}$, $\mu \geqslant 1$. Notons $u_r = q(t-2r)$, $u = (u_1, u_2, \ldots, u_\mu)$. $F(u_1, u_2, \ldots, u_\mu)$ est une fonction de $\mathbf{R}^{n\mu}$ dans $\mathbf{R}^n$ vérifiant la propriété suivante :

$$\big|F(u)\big| < 1, \quad F \in \mathrm{C}^1\big(\mathbf{R}^{n\mu}\big). \tag{5}$$

L'inégalité (5) montre que le système (4) est dissipatif.

3. Résultat (temps discret)

On désigne B^n la boule unité dans $\mathbf{R}^n$ de centre 0 et $B^{n,\mu}$ le produit
$B^{n,\mu} = B^n \times B^n \times \cdots \times B^n$ (μ fois).

THÉORÈME 1. – *Soient σ vérifiant la condition* (2) *et $\varepsilon > 0$. Alors, quels que soient les nombres positifs n et $\mu \in \mathbf{N}$ et la fonction F vérifiant la condition* (5)*, il existe un choix de paramétres $\mathcal{P} = \{N, \lambda, \theta, \mathbf{M}\}$ du système* (1) *tel que* :

(a) *ce système a la dynamique inertielle avec retard définie par* :

$$q(t) = F_{\mathcal{P}}\big(q(t-2), q(t-4), \ldots, q(t-2\mu)\big), \quad q \in B^{n,\mu}, \tag{6}$$

où $F_{\mathcal{P}}$ est un champ vectoriel qui satisfait aux conditions (5) *et, de plus, on a l'inégalité* :

$$|F - F_{\mathcal{P}}|_{\mathrm{C}^1(B^{n,\mu})} < \varepsilon; \tag{7}$$

(b) *les variables « cachées » q sont définies par une relation linéaire $q(t) = Ux(t)$, où U est une matrice constante.*

Remarque. – Etant donnés F et ε, les paramètres $\mathcal{P}$ et le champ $F_{\mathcal{P}}$ peuvent être définis par un algorithme complexe mais constructif.

4. Résultat (temps continu)

En temps continu, le réseau de Hopfield est donné par le système d'équations différentielles :

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j=1}^{N} K_{ij}\sigma(x_j) - b_i x_i + \theta_i. \tag{8}$$

Ici l'espace des paramètre $\mathcal{P}$ est $\mathcal{P} = \{N, \mathbf{K}, b, \theta\}$ où $\theta = (\theta_1, \ldots, \theta_N)$. Considérons un système dynamique dissipatif dans la boule $B^n \subset \mathbf{R}^n$ défini par les équations :

$$\frac{\mathrm{d}q}{\mathrm{d}t} = F(q), \quad q \in B^n, \ F \in \mathrm{C}^1\big(B^n\big). \tag{9}$$

THÉORÈME 2. – *Soient σ vérifiant la condition* (2) *et $\varepsilon > 0$. Alors, quels que soient les nombres positifs n et $\mu \in \mathbf{N}$ et le système dissipatif de la forme* (9)*, il existe un choix de paramètres $\mathcal{P}$ du système* (8) *tel que* :

(a) *La matrice $\mathbf{K}$ satisfait aux conditions* (3) *et les composantes des vecteurs b et θ ne prennent que deux valeurs distinctes ;*

(b) *le système* (8) *a la dynamique inertielle définie par* :

$$\frac{\mathrm{d}q}{\mathrm{d}t} = F_{\mathcal{P}}(q), \quad q \in B^n, \tag{10}$$

où $F_{\mathcal{P}}$ est un champ vectoriel satisfaisant à

$$|F - F_{\mathcal{P}}|_{C^1(B^n)} < \varepsilon. \tag{11}$$

Corollaires, remarques et applications. – Rappelons qu'un système dynamique $q \to F(q)$ est structurellement stable s'il existe un voisinage du champ F dans C^1 tel que tout champ de vecteurs de ce voisinage définisse un système topologiquement orbitalement équivalent au système initial, l'équivalence étant réalisée par un homéomorphisme voisin de l'identité.

C'est le cas par exemple pour les flots définis comme gradients d'une fonction de Morse d'une surface compacte, et dont les courbes invariantes des points – selle sont deux à deux disjointes.

La théorie dynamique a véritablement pris son essort quand on a démontré que les difféomorphismes de Thom–Anosov et le fer à cheval de Smale sont des exemples de systèmes dynamiques apparemment compliqués, mais stables par des petites perturbations C^1.

Ces dynamiques, et en général toute dynamique sans mémoire et structurellement stable, peuvent être engendrées par des réseaux de Hopfield, à une équivalence topologique près.

De plus nous remarquons que la structure de K de la forme (3) est une structure minimale pour obtenir toute dynamique structurellement stable.

En effet, considérons un réseau (1) avec une matrice synaptique $\mathbf{K}$ plus simple, par exemple telle que $K_{ij} \geqslant 0$ pour tous i, j. Alors, si ce système engendre une dynamique structurellement stable, elle ne peut être que périodique. En effet, ce système est proche d'un système monotone ayant comme attracteur un ensemble de cycles périodiques [6]. Par conséquent, des systèmes avec $K_{ij} \geqslant 0$ par exemple ne peuvent pas engendrer les fers à cheval de Smale.

Le résultat obtenu montre que, même les réseaux de neurones les plus simples tels que (1), peuvent réaliser des algorithmes différents de reconnaissance et classification [1,3].

5. Démonstration. Algorithme d'approximation

Supposons que $\mu = 1$ (le cas $\mu > 1$ peut être réduit à $\mu = 1$ par la substitution $Q(t) = (q(t), \; q(t-2), \ldots, q(t-2\mu))^t$). On va opérer en quatre étapes.

Etape 1 : *Approximation particulière de F.* – Nous suivons les travaux [9,8]. Définissons une famille de champs vectoriels sur $\mathbf{R}^n$, dépendant d'un paramètre $\mathcal{P}$, par

$$F_{\mathcal{P}}(u) = \sum_{i=1}^{N_1} B_i \sigma(\mathbf{A}_i u + \theta_i), \tag{12}$$

où $\mathcal{P} = \{\mathbf{A}, \mathbf{B}, \theta, N_1\}$, $\mathbf{A}$ et $\mathbf{B}$ des matrices de dimensions $N_1 \times n$ et $n \times N_1$, et θ un vecteur de $\mathbf{R}^{N_1}$.

Le champs de vecteurs $F(u)$ (qui définit la dynamique (4)) peut être approché, dans la boule B^n, par une somme $F_{\mathcal{P}}$ de la forme (12) :

$$|F - F_{\mathcal{P}}|_{C^1(B^n)} < \varepsilon/2. \tag{13}$$

La démonstration de (13) est basée sur l'identité de Calderón [9,8], ce fait est proche de résultats connus en théorie des réseaux en couches. Voir, par exemple, [5,4].

Etape 2 : *Définition des variables cachées.* – Nous suivons le raisonnement donné dans [9]. Définissons un réseau par les relations suivantes :

$$x_i(t) = \sigma\bigl(\mathbf{P}_i x(t-2) + \theta_i\bigr), \tag{14}$$

où la matrice $\mathbf{P}$ de dimensions $N_1 \times N_1$ est définie par $\mathbf{P} = \mathbf{AB}$. Les matrices $\mathbf{A}$ et $\mathbf{B}$ et le vecteur θ sont déterminés à l'étape précédente. Alors, on introduit les variables « cachées » q par $q = \mathbf{B}x$, $q \in \mathbf{R}^n$ et on voit, pour $t > 2$, que la dynamique (14) entraîne

$$q(t) = F_{\mathcal{P}}\bigl(q(t-2)\bigr). \tag{15}$$

Par conséquent, il ne reste plus qu'à réduire la dynamique (1) à la dynamique (14), à $\varepsilon/2$ près.

Etape 3 : *Organisation spéciale du réseau.* – Nous considérons un réseau d'une structure particulière avec deux types de neurones : les x-neurones et les y-neurones. On désigne les états des x-neurones par x_i, $i = 1, 2, \ldots, N_1$, et les états des y-neurones par y_k, $k = 1, 2, \ldots, N_2$. Alors $N = N_1 + N_2$. Soient les matrices $\mathbf{R}$ et $\mathbf{S}$ de dimensions respectives $N_1 \times N_2$ et $N_2 \times N_1$. Signalons que la matrice $\mathbf{S}$ décrit une action des x-neurones sur les y-neurones et la matrice $\mathbf{R}$ définit une action réciproque.

Toutes des autres connexions sont absentes. Les colonnes de ces matrices sont désignées par $\mathbf{R}_i$ et $\mathbf{S}_k$. Nous supposons que ces matrices vérifient la condition (3).

Alors, nous obtenons les relations suivantes :

$$x_i(t+1) = \sigma\big(\mathbf{R}_i y(t) + \theta\big), \qquad y_k(t+1) = \sigma\big(\mathbf{S}_k x(t) + \theta\big), \tag{16}$$

où $i = 1, \ldots, N_1$, $k = 1, 2, \ldots, N_2$. On déduit de cette structure que $\mathbf{K}$ vérifie (3) si les éléments de $\mathbf{S}$ et $\mathbf{R}$ prennent les valeurs λ, $-\lambda$ ou 0.

Fixons N_1 et supposons N_2 est grand. De plus, soient a une constante et $\lambda = aN_2^{-1/2}$ (ainsi λ est petit). Par ailleurs, soient b une constante et $\theta = b\lambda$. Puisque $|x_i(t)| < C$, nous en déduisons que

$$x_i(t) = \sigma\big(\mathbf{K}_i x_j(t-2) + \bar{\theta}_i\big) + \lambda g_i\big(x(t-2)\big), \tag{17}$$

où

$$\mathbf{K} = \mathbf{RS}, \qquad \bar{\theta}_i = \theta\left(1 + \sum_{k=1}^{N_2} R_{ik}\right), \tag{18}$$

et puisque $|g_i|_{C^1} < c$, $\lambda g_i(x)$ est une petite correction.

Etape 4. – Soient deux nombres N_1 et n, une matrice $\widetilde{\mathbf{P}}$ de la forme $\widetilde{\mathbf{P}} = \mathbf{AB}$, un vecteur $\tilde{\theta}$ et $\delta > 0$ (voir étape 2), alors on peut déterminer des nombres a, b et des matrices $\mathbf{R}$, $\mathbf{S}$ telles que pour N_2 suffisamment grands on ait

$$\big|P_{ij} - K_{ij}(\mathbf{R}, \mathbf{S})\big| < \delta, \quad \big|\tilde{\theta}_i - \bar{\theta}_i(\mathbf{R})\big| < \delta, \tag{19}$$

où $\mathbf{K}$ et $\bar{\theta}_i$ sont définis par (18).

La démonstration repose sur une idée physique (voir [2]) et des arguments de probabilité. Nous décomposons les neurones en « populations ». Soit $X_1, X_2, \ldots, X_{n+1}$ une décomposition de l'ensemble $X = \{1, 2, \ldots, N_2\}$ en sous-ensembles disjoints. Le nombre d'éléments dans X_i est $M_i = \mathrm{O}(N_2)$. Si $k \in X_i$, nous posons $R_{ik} = \lambda$ ou $R_{ik} = -\lambda$ avec les probabilités respectives $\rho_j, 1 - \rho_j$ et $S_{jk} = \lambda$ ou $-\lambda$ avec probabilités respectives $\tilde{\rho}_j, 1 - \tilde{\rho}_j$. Alors la moyenne $N_2^{-1} \sum_k R_{ik} S_{jk}$ peut être considérée comme une espérance mathématique. On peut démontrer que l'on obtient ainsi l'approximation (19) pour $N_2 \to \infty$.

Pour le cas continu, la démonstration est analogue, à quelques details près, et elle utilise des idées connues [7].

Références bibliographiques

[1] L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel, Axioms and fundamental equations of image processing, Arch. Rational Mech. Anal. 16 (1993) 200–257.

[2] N.O. Brunel, O. Truillier, Plasticity of directional place fields in a model of rodent CA3, Hippocampus 8 (1998) 651–665.

[3] V. Caselles, B. Coll, J.M. Morel, Partial differential equations and image processing, Séminaire Équations aux Dérivées Partielles, École polytechnique, Exp. XXI (1995–1996) XXI-1–XXI-30.

[4] K. Funahashi, Y. Nakamura, The approximation of dynamical systems by continuous time reccurent neural networks, Neural Networks 6 (1993) 801–806.

[5] K. Hornik, M. Stinchcombe, H. White, Multilayered feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.

[6] P. Polácik, T. Terescák, Convergence to cycles as a typical asymptotic behaviour in smooth strongly monotone dicrete-time dynamical systems, Arch. Rational Mech. Anal. 116 (1991) 339–360.

[7] R. Temam, Infinite Dimensional Dynamical Systems in Mechanics and Physics, Springer, 1988.

[8] S. Vakulenko, Dissipative systems generating any structurally stable chaos, Adv. Differential Equations 5 (2000) 1139–1178.

[9] S. Vakulenko, P. Gordon, Neural networks with prescribed large time behaviour, J. Phys. A 31 (1998) 9555–9570.