

Signaux et systèmes/*Signal and Systems*

Maximisation de l'entropie informationnelle dans la transmission par une non-linéarité neuronale

François CHAPEAU-BLONDEAU

Résumé – Pour une unité continue munie d'une non-linéarité typique d'un neurone (monotone croissante à seuil et saturation), nous déterminons comment l'entropie informationnelle est transformée dans la transmission entrée-sortie. Un principe de maximisation de l'entropie de sortie est introduit, que l'on associe à une représentation optimale de l'information. Avec la non-linéarité neuronale, les signaux d'entrée gaussiens permettent quasiment d'atteindre le maximum absolu de l'entropie de sortie, ce qui suggère une adaptation optimale de cette non-linéarité à l'environnement physique. Pour des signaux quelconques, la maximisation adaptative de l'entropie de sortie conduit à un algorithme de plasticité synaptique de caractère hebbien.

Information entropy maximization in the transmission by a neuron nonlinearity

Abstract – For a continuous unit endowed with a neuron-like nonlinearity (monotonically increasing with threshold and saturation), the transformation of the information entropy in the input-output transmission is derived. A principle of maximization of the output entropy is introduced, that we associate to an optimal representation of information. With the neural nonlinearity, Gaussian input signals lead nearly to the absolute maximum of the output entropy, suggesting an optimal adaptation of this nonlinearity to the physical environment. For arbitrary signals, adaptive maximization of the output entropy leads to a synaptic plasticity algorithm of Hebbian character.

Abridged English Version – Application of information theory to the analysis of information transfer in neural networks, has mainly dealt with noise-free discrete (binary) nonlinear neurons or noisy continuous linear neurons (Taylor, 1993). These assumptions allow the definition of the information transmitted by the neuron from input-output conditional entropy (Linsker, 1990; Atick, 1992), and then the determination of conditions for its maximization. We apply here information theory to noise-free continuous nonlinear neurons. In such conditions, the transmitted information classically defined goes to infinity, and cannot serve to characterize the system. We examine the characterization afforded by the consideration of output entropy rather than transmitted information, and show its usefulness for transcoding of sensory signals.

A model is considered for a sensory neuron excited by a pool of sources or sensors with sufficient homogeneity to justify the definition of a single input signal x . Typically, x gives an image of a physical quantity of the environment, uniform on the sensors feeding the neuron, and that we endow with the probability density function (pdf) $p_x(x)$. The neuron produces a nonlinear continuous output response y of pdf $p_y(y)$, that is interpreted as a firing rate. Information entropies for x and y are defined according to equation (1).

The output y being limited to a finite range $[y_{\min}, y_{\max}]$, the output entropy $H(y)$ reaches its maximum H_{\max} of equation (4) for the uniform output pdf of equation (3). Maximizing the entropy of the output y is, in some sense, an optimal use of the finite interval of representation $[y_{\min}, y_{\max}]$.

A typical neuronal input-output nonlinearity (monotonically increasing with threshold and saturation) can be modelled with the sigmoid φ_1 of equation (5), where the threshold θ and the synaptic efficacy w (or equivalent synaptic efficacy for a homogeneous bundle of pathways carrying the same input signal) are both adaptable parameters.

Note présentée par Gérard TOULOUSE.

When $y = \varphi_1(x)$, the input pdf that maximizes $H(y)$ is given by equation (6), which is almost a Gaussian pdf. A neuron nonlinearity is thus efficient for the transduction of Gaussian signals that are ubiquitous in the physical environment.

Other nonlinear transfer functions are frequently used for information processing, such as φ_2 of equation (7) (Park, 1991). With nonlinearity φ_2 , the maximization of $H(y)$ occurs for the bimodal input pdf of equation (8), less frequent in the physical environment.

In the input-output transmission $y = \varphi_1(x)$, we establish that the transformation of the signal entropies occurs according to equations (10), (11). With arbitrary input, the range of $H(x)$ is arbitrary. In contrast, $H(y)$ is bounded by H_{\max} of equation (4) because of the confinement of y to a finite interval. H_{trans} expresses the change of entropy in the transcoding that maps to a finite interval $[y_{\min}, y_{\max}]$ an input x that can vary in an arbitrary interval. H_{trans} can thus be positive or negative, whether the transcoding implements respectively a nonlinear "expansion" or "contraction".

For an arbitrary input signal x it is considered that the most efficient output representation is the one that maximizes entropy $H(y)$. Maximization of $H(y)$, or equivalently of $H_{\text{trans}}(\theta, w)$, can be performed by adaptation of the parameters θ and w through a gradient ascent expressed in equations (12)-(15). It is a remarkable feature that the synaptic adaptation law (15), deduced from the maximization of entropy $H(y)$, displays a Hebbian form. Similar antihebbian laws have already been obtained as a consequence of general information processing principles, for instance to decorrelate output signals in a network of noisy linear neurons with lateral inhibition (Plumbley, 1993). Such properties demonstrate the usefulness of general information processing principles for the understanding of neural systems.

The figure illustrates the evolution of $p_y(y)$ toward maximum entropy during adaptive changes to θ and w according to equations (12)-(14) with $\alpha = 0.5$.

1. INTRODUCTION. – L'application de la théorie statistique de l'information pour l'analyse du transfert d'information dans les réseaux de neurones, a porté principalement sur des neurones non linéaires à sortie discrète (binaire) ou sur des neurones linéaires à sortie continue bruitée (Taylor, 1993). Ces hypothèses rendent possible la définition de l'information transmise par les neurones, à partir de l'entropie conditionnelle de la sortie par rapport à l'entrée (Linsker, 1990; Atick, 1992), puis la détermination de conditions pour sa maximisation.

On peut chercher à appliquer la théorie de l'information à des modèles de neurones non linéaires à sortie continue, où le bruit peut être choisi arbitrairement petit. Dans ces conditions, l'information transmise définie classiquement tend vers l'infini, et ne permet plus de caractériser le système. Pour cette catégorie de neurones, nous considérons ici l'entropie de sortie au lieu de l'information transmise, et examinons le type de caractérisation qu'elle autorise. Nous montrons les implications d'un principe de maximisation de l'entropie de sortie d'un neurone, et son intérêt pour le transcodage de signaux sensoriels, lorsque l'on cherche les conditions optimisant la représentation d'une grandeur sur une sortie neuronale non-linéaire variant dans un intervalle fini.

2. SIGNAUX ET ENTROPIE. – On considère la situation d'un neurone sensoriel, alimenté en entrée par un ensemble de sources ou capteurs suffisamment homogène pour permettre de définir un unique signal d'entrée x . Typiquement, x donne l'image d'une grandeur physique présente dans l'environnement, et uniforme sur les capteurs alimentant le neurone examiné.

On considère x continu, muni d'une densité de probabilité $p_x(x)$. En réponse, le neurone produit un signal de sortie continu y de densité de probabilité $p_y(y)$, qui peut s'interpréter comme une fréquence de décharge, et qui est une fonction généralement non linéaire de l'entrée x . On définit l'entropie informationnelle $H(x)$ de x comme (Shannon, 1948):

$$(1) \quad H(x) = - \int_{-\infty}^{+\infty} p_x(x) \log_2 p_x(x) dx,$$

et une définition analogue pour l'entropie $H(y)$ de y .

3. LOI EN SORTIE QUI MAXIMISE L'ENTROPIE DE SORTIE. – Si l'on interprète la sortie y comme une fréquence de décharge neuronale, il est naturel de restreindre les variations de y à un intervalle fini. Pour une variable aléatoire $y \in [y_{\min}, y_{\max}]$,

$$(2) \quad \max \text{ de } H(y) = - \int_{y_{\min}}^{y_{\max}} p_y(y) \log_2 p_y(y) dy \quad \text{sous} \quad \int_{y_{\min}}^{y_{\max}} p_y(y) dy = 1,$$

s'obtient pour la loi uniforme $p_y(y) = p_y^{\text{opt}}(y)$ telle que:

$$(3) \quad p_y^{\text{opt}}(y) = \frac{1}{y_{\max} - y_{\min}} \quad \text{pour } y \in [y_{\min}, y_{\max}], \quad \text{et} \quad p_y^{\text{opt}}(y) = 0 \quad \text{sinon.}$$

Le maximum de l'entropie $H(y)$ atteint vaut alors:

$$(4) \quad H_{\max} = \log_2 (y_{\max} - y_{\min}).$$

Maximiser l'entropie du signal de sortie y constitue, en un certain sens, une utilisation optimale de l'intervalle fini de représentation $[y_{\min}, y_{\max}]$.

4. LOI EN ENTRÉE QUI MAXIMISE L'ENTROPIE DE SORTIE. – Une fonction non linéaire typique $y = \varphi_1(x)$ qui constitue un modèle intéressant pour le transfert entrée-sortie d'un neurone, est constituée par la fonction sigmoïde d'expression:

$$(5) \quad \varphi_1(x) = \frac{1}{1 + \exp[-\beta(wx - \theta)]}.$$

Ici, la pente $\beta > 0$ de la sigmoïde est fixée; le seuil θ , ainsi que l'efficacité w de la synapse (ou synapse équivalente pour un ensemble homogène de voies véhiculant le même signal d'entrée) appliquant l'entrée x au neurone, sont considérés adaptables. La sigmoïde de l'équation (5) modélise bien, au moins de manière qualitative, trois propriétés importantes du transfert entrée-sortie d'un neurone: le caractère monotone croissant de la réponse, l'existence d'un seuil et d'une saturation.

En présence de ce type de non-linéarité neuronale typique φ_1 , il est intéressant de rechercher la loi $p_x^{\text{opt}}(x)$ du signal d'entrée x qui permet d'atteindre le maximum H_{\max} de l'entropie en sortie donnée par l'équation (4). Comme $y = \varphi_1(x)$ donne $y \in [y_{\min} = 0, y_{\max} = 1]$, cela revient, conformément à l'équation (3), à chercher à obtenir en sortie $p_y(y) = p_y^{\text{opt}}(y) = 1$ pour $y \in [0, 1]$ et $p_y(y) = p_y^{\text{opt}}(y) = 0$ sinon. La conservation des probabilités dans la transformation par φ_1 donne $p_x^{\text{opt}}(x) = p_y^{\text{opt}}(y) dy/dx$, d'où il vient:

$$(6) \quad p_x^{\text{opt}}(x) = \frac{\beta w}{4 \cosh^2 [(\beta/2)(wx - \theta)]}.$$

Il apparaît que la loi optimale de l'équation (6) est en fait très proche d'une loi gaussienne. Il est donc remarquable de constater que pour une non-linéarité typique d'un

neurone, le maximum de l'entropie en sortie puisse être quasiment atteint en présence de signaux d'entrée gaussiens. Les signaux gaussiens se rencontrent très fréquemment dans l'environnement physique. Une non-linéarité neuronale typique semble donc spécialement bien adaptée à leur transcodage efficace. Peut-être y a-t'il ici un principe de développement qui a conduit les neurones à favoriser un certain type de réponse non-linéaire, bien adaptée à leur environnement physique.

D'autres types de fonctions de transfert non linéaires sont utilisées fréquemment pour le traitement de l'information. C'est le cas des fonctions de transfert gaussiennes (Park, 1991) de la forme (avec $\beta > 0$) :

$$(7) \quad \varphi_2(x) = \exp[-\beta(wx - \theta)^2].$$

Pour ce type de non-linéarité, le maximum d'entropie en sortie est atteint avec la loi d'entrée :

$$(8) \quad p_x^{\text{opt}}(x) = \beta|wx - \theta| \exp[-\beta(wx - \theta)^2].$$

La loi optimale de l'équation (8) adaptée à la non-linéarité φ_2 est une loi bimodale moins fréquemment rencontrée que la loi gaussienne. La non-linéarité φ_2 est donc adaptée au transcodage de signaux d'un type moins répandu dans l'environnement physique.

5. TRANSFORMATION DE L'ENTROPIE DANS LA TRANSMISSION ENTRÉE-SORTIE. – La transmission par la fonction de transfert $y = \varphi_1(x)$ s'accompagne de la relation $p_x(x) dx = p_y(y) dy$, ou encore $p_x(x) = p_y(y) \varphi_1'(x)$. On a donc pour l'entropie d'entrée :

$$(9) \quad - \int_{-\infty}^{+\infty} p_x(x) \log_2 p_x(x) dx = - \int_{-\infty}^{+\infty} p_y(y) \log_2 [p_y(y) \varphi_1'(x)] dy.$$

En développant le logarithme du produit, on déduit que la transformation de l'entropie dans la transmission entrée-sortie peut s'exprimer par :

$$(10) \quad H(y) = H(x) + H_{\text{trans}},$$

où H_{trans} , qui représente la variation d'entropie dans la transmission, vaut :

$$(11) \quad H_{\text{trans}} = \int_{-\infty}^{+\infty} p_x(x) \log_2 \varphi_1'(x) dx.$$

L'entropie d'entrée $H(x)$ peut être arbitrairement grande ou petite, selon en particulier la variance de x . Ainsi pour une entrée x gaussienne de variance σ_x^2 , on a $H(x) = \log_2(\sigma_x \sqrt{2\pi e})$. Par contre, l'entropie de sortie $H(y)$ vaut au maximum H_{max} donné par l'équation (4), car la sortie y est confinée dans un intervalle fini. La variation d'entropie H_{trans} de l'équation (11) peut donc être positive ou négative. Elle traduit l'action du transcodage qui ramène à un intervalle fini de représentation $[y_{\text{min}}, y_{\text{max}}]$, une entrée x pouvant varier dans un intervalle arbitraire. Cette action peut être réalisée par « dilatation » ou « contraction » non linéaires, associées respectivement à $H_{\text{trans}} > 0$ et $H_{\text{trans}} < 0$.

6. MAXIMISATION ADAPTATIVE DE L'ENTROPIE DE SORTIE. – Quelle que soit l'entrée x , on considère ici que la représentation de sortie la plus efficace est celle qui maximise l'entropie $H(y)$. En présence d'une entrée x de loi $p_x(x)$ fixée, maximiser $H(y)$ équivaut à maximiser H_{trans} . Pour la fonction de transfert φ_1 de l'équation (5) avec β donné, on

cherche alors à obtenir la valeur des paramètres adaptables θ et w qui maximise H_{trans} . Le calcul explicite des dérivées premières et secondes montre que $H_{\text{trans}}(\theta, w)$ est une fonction concave à un seul maximum. La maximisation de $H_{\text{trans}}(\theta, w)$ peut être réalisée par une ascension selon le gradient de H_{trans} , où l'on adapte les paramètres suivant un schéma dynamique de la forme (avec $\alpha > 0$) :

$$(12) \quad \frac{d\theta}{dt} = \alpha \frac{\partial H_{\text{trans}}}{\partial \theta}, \quad \text{et} \quad \frac{dw}{dt} = \alpha \frac{\partial H_{\text{trans}}}{\partial w}.$$

Le calcul des composantes du gradient donne :

$$(13) \quad \frac{\partial H_{\text{trans}}}{\partial \theta} = 2\beta \left(\bar{y} - \frac{1}{2} \right),$$

$$(14) \quad \frac{\partial H_{\text{trans}}}{\partial w} = \frac{1}{w} + \beta (\bar{x} - 2\bar{xy}),$$

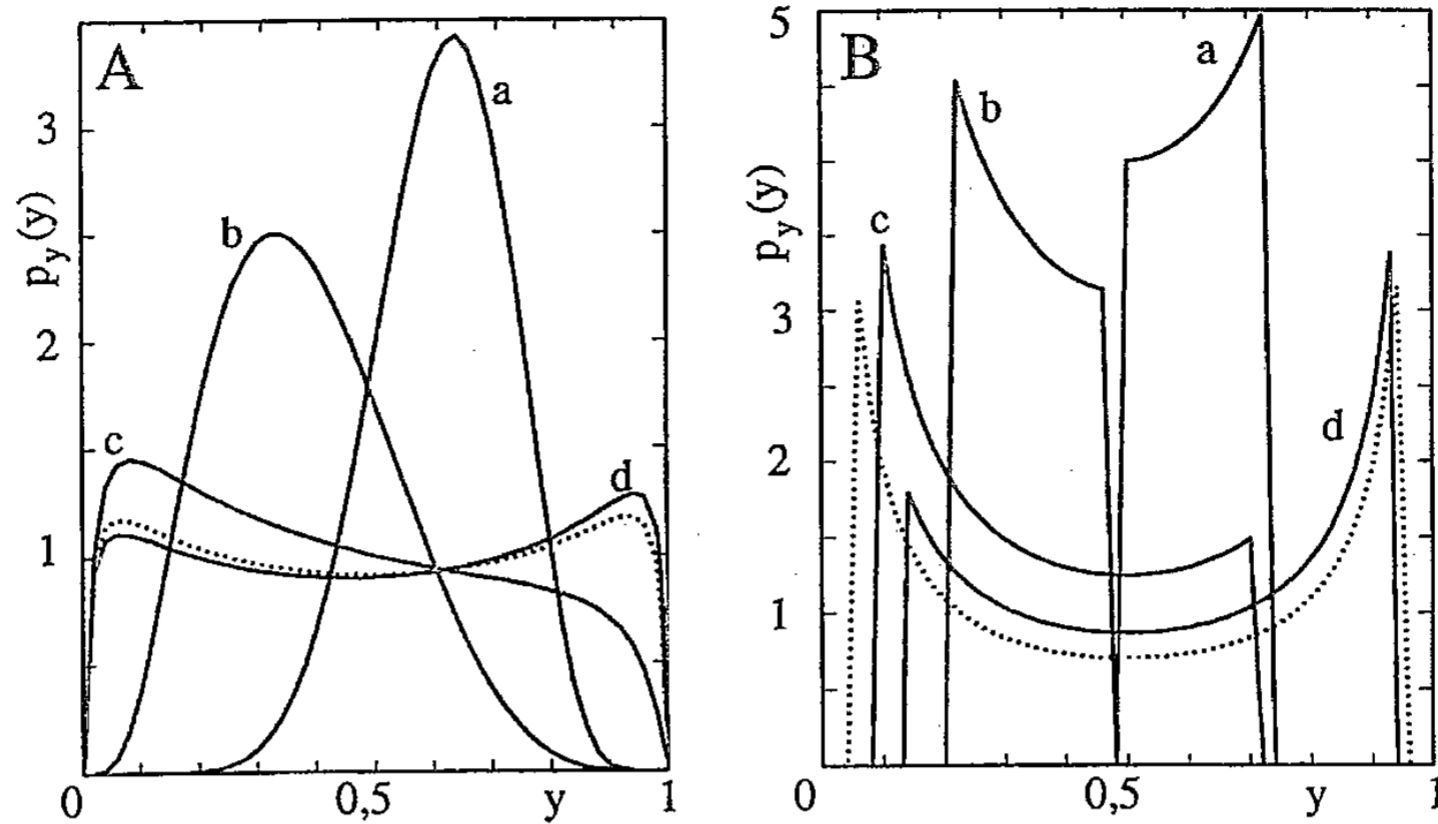
où pour une fonction $u(x)$ quelconque de x , on introduit la moyenne $\bar{u} = \int_{-\infty}^{+\infty} u(x) p_x(x) dx$.

D'après l'équation (13), le maximum de H_{trans} est associé à $\bar{y} = 1/2$. Ainsi, au maximum de l'entropie de sortie $H(y)$ la valeur moyenne de la sortie y est située au milieu de l'intervalle de représentation [$y_{\text{min}} = 0, y_{\text{max}} = 1$], ce qui est une propriété naturelle, qui découle ici du principe de maximisation de l'entropie de sortie. Si l'on introduit les fluctuations autour des moyennes $\tilde{x} = x - \bar{x}$ et $\tilde{y} = y - 1/2$, l'adaptation de l'efficacité synaptique w prend la forme :

$$(15) \quad \frac{dw}{dt} = \alpha \left(\frac{1}{w} - 2\beta \overline{\tilde{x}\tilde{y}} \right).$$

Il est intéressant de remarquer que l'équation (15) peut s'interpréter comme une loi de caractère hebbien (ou antihebbien), puisque la variation de l'efficacité synaptique dépend à la fois des signaux d'entrée et de sortie du neurone, par l'intermédiaire d'une moyenne de leur produit $\tilde{x}\tilde{y}$. Il est donc frappant de constater que le principe abstrait de maximisation de l'entropie de sortie du neurone, réalisée par une méthode formelle d'ascension de gradient, conduit à une loi de plasticité synaptique de forme hebbienne. Ce type de loi antihebbienne a déjà été rencontré comme conséquence de principes généraux de traitement de l'information, par exemple pour réaliser la décorrélation de signaux de sortie dans un réseau de neurones linéaires bruités à inhibition latérale (Plumbley, 1993). De telles propriétés confirment l'utilité que peut présenter l'introduction de principes généraux de traitement de l'information pour la compréhension des systèmes neuronaux.

La figure illustre l'évolution de $p_y(y)$ vers le maximum d'entropie (densité uniforme), lors de l'adaptation des paramètres θ et w par (12)-(14) avec $\alpha = 0,5$, en présence de différents signaux d'entrée numériques $x(t)$ au pas $\Delta t = 1$. L'estimation des moyennes des équations (13), (14) est réalisée dans le domaine temporel par un filtre linéaire passe-bas du 1^{er} ordre de constante de temps $\tau = 25$. On observe une convergence relativement rapide vers le maximum d'entropie, facilitée par un H_{trans} toujours concave, à l'intérieur du compromis usuel en traitement adaptatif entre vitesse de convergence et erreur résiduelle.



Évolution de $p_y(y)$ vers le maximum d'entropie lors de l'adaptation de θ et w par (12)-(14). A) Avec $p_x(x)$ gaussienne de $\bar{x} = 1,5$ et $\sigma_x = 0,5$: (a) état initial à $t = 0$ avec $\theta = 1$ et $w = 1$, $H(y) = -1,125$; (b) à $t = 50$, $H(y) = -0,728$; (c) à $t = 100$, $H(y) = -0,067$; (d) à $t = 200$, $H(y) = -0,015$. En pointillés figure l'optimum théorique qui réalise le maximum absolu $H(y) = -0,013$ pour une entrée x gaussienne quelconque. B) Avec $p_x(x)$ uniforme sur $[1, 2]$: (a) état initial à $t = 0$ avec $\theta = 1$ et $w = 1$, $H(y) = -2,117$; (b) à $t = 50$, $H(y) = -1,871$; (c) à $t = 100$, $H(y) = -0,757$; (d) à $t = 200$, $H(y) = -0,410$. En pointillés figure l'optimum théorique qui réalise le maximum absolu $H(y) = -0,320$ pour une entrée x uniforme quelconque.

Evolution of $p_y(y)$ toward maximum entropy during adaptation of θ and w through (12)-(14). A) With Gaussian $p_x(x)$ of $\bar{x} = 1.5$ and $\sigma_x = 0.5$: (a) initial state at $t = 0$ with $\theta = 1$ and $w = 1$, $H(y) = -1.125$; (b) at $t = 50$, $H(y) = -0.728$; (c) at $t = 100$, $H(y) = -0.067$; (d) at $t = 200$, $H(y) = -0.015$. The dotted curve is the theoretical optimum that realizes the absolute maximum $H(y) = -0.013$ for an arbitrary Gaussian input x . B) With $p_x(x)$ uniform over $[1, 2]$: (a) initial state at $t = 0$ with $\theta = 1$ and $w = 1$, $H(y) = -2.117$; (b) at $t = 50$, $H(y) = -1.871$; (c) at $t = 100$, $H(y) = -0.757$; (d) at $t = 200$, $H(y) = -0.410$. The dotted curve is the theoretical optimum that realizes the absolute maximum $H(y) = -0.320$ for an arbitrary uniform input x .

Note remise le 10 janvier 1994, acceptée après révision le 13 juin 1994.

RÉFÉRENCES BIBLIOGRAPHIQUES

- J. J. ATICK, Could information theory provide an ecological theory of sensory processing?, *Network: Computation in Neural Systems*, 3, 1992, p. 213-251.
 R. LINSKER, Perceptual neural organization: Some approaches based on network models and information theory, *Annual Review of Neuroscience*, 13, 1990, p. 257-281.
 J. PARK et I. W. SANDBERG, Universal approximation using radial-basis function networks, *Neural Computation*, 3, 1991, p. 246-257.
 M. D. PLUMBLEY, Efficient information transfer and anti-hebbian neural networks, *Neural Networks*, 6, 1993, p. 823-833.
 C. E. SHANNON, A mathematical theory of communication, *Bell System Technical Journal*, 27, 1948, p. 623-656.
 J. G. TAYLOR et M. D. PLUMBLEY, Information theory and neural networks, In J. G. TAYLOR éd., *Mathematical Approaches to Neural Networks*, Elsevier, Amsterdam, 1993.

Institut de Biologie Théorique, Université d'Angers, 10, rue André Boquet, 49100 Angers, France.