

Théorème d'approximation pour les neurones hypersphériques

Julien DE SAINT ANGEL¹ Christophe SAINT-JEAN¹

¹Laboratoire Mathématiques Image Application (MIA),
Faculté des Sciences et Technologies, Av. Henri Becquerel, 17000 La Rochelle

Résumé – Cet article établit un théorème d'approximation pour un réseau de neurones à une couche hypersphérique. Pour cela, nous écartons de la démonstration classique et utilisons un théorème de Schwartz basé sur une combinaison linéaire de translations. Ce théorème d'approximation n'est pas universel dans le sens où les fonctions à approcher comme les activations doivent être à support compact. Également, nous établissons le lien entre un réseau à couche hypersphérique et un réseau à de fonctions à bases radiales. Des expériences ont permis de vérifier le théorème et de comparer les approximations par couche sphérique et par couche dense. Par ailleurs, celles-ci ont également permises d'établir expérimentalement que les conditions de compacité sur les fonctions d'activations sont suffisantes mais non nécessaire. Ceci ouvre la voie à une approche plus classique du théorème d'approximation universelle pour alléger les hypothèses sur la nature des fonctions que l'on peut approximer.

Abstract – This paper establishes an approximation theorem for a hyperspherical single layer neural network. For this, we depart from the classical proof and use a Schwartz theorem based on a linear combination of translations. This approximation theorem is not universal in the sense that the functions to be approximated like the activations must be compactly supported. also establish the link between a hyperspherical layer network and a radial basis functions network. Experiments have allowed us to verify the theorem and to compare the spherical layer and dense layer approximations. Moreover, these experiments also allowed to establish experimentally that the compactness conditions on the activation functions are sufficient but not necessary. This opens the way to a more classical approach of the universal approximation theorem to alleviate the assumptions on the nature of the functions that can be approximated.

1 Introduction

Dans les réseaux de neurones, une couche dense est définie par $\mathbf{W}\mathbf{x} + \mathbf{b}$, où \mathbf{W} est une matrice $m \times n$, \mathbf{x} le vecteur d'entrée de taille n et \mathbf{b} un vecteur de taille m . Les éléments \mathbf{W} et \mathbf{b} correspondent aux paramètres de la couche dense et définissent géométriquement m hyperplans par le système d'équations $\mathbf{W}\mathbf{x} + \mathbf{b} = 0$.

Dans une couche hypersphérique qui utilise l'algèbre conforme, les m hyperplans deviennent des m hypersphères ; l'équation $w_j \cdot x + b_j = 0$ devient le produit conforme $\tilde{s}_j \cdot \tilde{x} = 0$ qui sera explicité dans la partie 2. Ce type de couche proposé par Banarar *et al.* [1] a été essentiellement utilisé dans un contexte de classification ou de filtre convolutif [2]. En particulier, les auteurs avaient constaté que pour une précision équivalente, une couche hypersphérique nécessitait moins de paramètres qu'une couche de type dense lorsque les données d'entrée sont en grande dimension.

Dans cet article, on s'intéresse à l'utilisation d'une couche hypersphérique dans le contexte de la régression. Naturellement, on se pose la question pour ce type de couche, si l'on peut approximer, au sens de la norme uniforme, toute fonction $f \in C^0(K)$ (continue sur un compact de \mathbb{R}^n) par une combinaison linéaire de m sorties de neurones hypersphériques. Ceci correspond à la version du théorème d'approximation universelle dans le cas d'une couche dense cachée à largeur m arbitraire. Dans [5], A. Pinkus effectue un état de l'art complet sur la question de densité qui consiste à montrer que l'ensemble $\mathcal{M}(\sigma) := \text{vect}\{\sigma(\mathbf{w} \cdot \mathbf{x} - b) : b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n\}$ est dense dans $C^0(K)$ avec K un compact de \mathbb{R}^n .

La question principale du papier est d'étudier les conditions dans lesquelles une fonction peut être approximée par un réseau de neurones à une seule couche hypersphérique de largeur arbitraire. Les contributions de cet article se résument par l'établissement d'un théorème d'approximation pour un réseau à une couche hypersphérique puis, la vérification expérimental ce théorème en identifiant des fonctions d'activation susceptibles de mieux fonctionner. Nous comparons également les résultats d'approximation avec ceux obtenus par un réseau à une couche dense.

L'article se décompose en trois parties : le cadre mathématique permettant de définir un neurone et une couche hypersphérique, un théorème d'approximation relatif à notre cas d'étude, une validation expérimentale du théorème qui offre des perspectives.

2 Neurone et couche hypersphériques

Les neurones et couches hypersphériques sont définis dans le cadre de l'algèbre géométrique conforme $\mathcal{G}(n+1, 1)$ [3] où n sera la dimension de l'entrée de la couche. Cette algèbre est construite à partir de \mathbb{R}^n (base $\{e_1, \dots, e_n\}$) augmenté de l'espace de Minkowski (base $\{e_+, e_-\}$). Dans ce dernier, une nouvelle base $\{e_0, e_\infty\}$ est définie telle que e_0 (resp. e_∞) représente le point homogène représentant l'origine de \mathbb{R}^n (resp. le point homogène à l'infini) (voir Eq. (2.28) dans [3]). Les règles du produit interne noté \cdot entre les éléments de la base sont les suivantes :

$$\begin{cases} e_i \cdot e_j = \delta_{ij}, & e_0 \cdot e_i = e_\infty \cdot e_i = 0 \text{ pour } i, j = 1..n \\ e_0 \cdot e_0 = e_\infty \cdot e_\infty = 0, & e_0 \cdot e_\infty = -1 \end{cases}$$

Dans l'algèbre $\mathcal{G}(n+1, 1)$ de dimension 2^{n+2} , l'ensemble des vecteurs forment un sous-espace de dimension $n+2$ (base $\{e_1, \dots, e_n, e_0, e_\infty\}$). Les points et les sphères de l'espace euclidien \mathbb{R}^n peuvent être représentés de manière unique par des vecteurs particuliers :

- Un point $\mathbf{x} \in \mathbb{R}^n$ est représenté par le vecteur \tilde{x} :

$$\tilde{x} = \mathbf{x} + \frac{1}{2}\mathbf{x}^2 e_\infty + e_0 = \mathbf{x} + \frac{1}{2}\|\mathbf{x}\|^2 e_\infty + e_0$$

où $\|\cdot\|$ est la norme euclidienne usuelle. Ce plongement qui est la première opération effectuée sur l'entrée \mathbf{x} de la couche, vérifie $\tilde{x}^2 = \tilde{x} \cdot \tilde{x} = 0$.

- Une hypersphère de centre $\mathbf{c} \in \mathbb{R}^n$ et de rayon $\rho \in \mathbb{R}^+$ sera représentée par le vecteur \tilde{s} :

$$\tilde{s} = \tilde{c} - \frac{\rho^2}{2} e_\infty = \mathbf{c} + \frac{1}{2}(\|\mathbf{c}\|^2 - \rho^2) e_\infty + e_0$$

pour lequel on observe que $\tilde{s}^2 = \tilde{s} \cdot \tilde{s} = \rho^2$. La coordonnée devant e_0 étant égale à 1 (équivalent à $\tilde{s} \cdot e_\infty = -1$), le vecteur \tilde{s} à $n+1$ paramètres libres comme le couple (w, b) d'un neurone classique.

On complète notre description par l'action de \tilde{s} sur \tilde{x} . La sortie sans activation d'un neurone hypersphérique est donnée par le produit intérieur :

$$\tilde{s} \cdot \tilde{x} = -\frac{1}{2} \sum_{i=1}^n \left[\tilde{x}_i^2 - 2\tilde{s}_i \tilde{x}_i + \tilde{s}_i^2 - \frac{\tilde{s}^2}{n} \right]. \quad (1)$$

où $\tilde{x}_i := \tilde{x} \cdot e_i$ et $\tilde{s}_i := \tilde{s} \cdot e_i$. Cette expression est un polynôme de degré 2 dans chacune des n directions. La figure 1 illustre la valeur de $\tilde{s} \cdot \tilde{x}$ (en vert) pour $n=1, c=0, \rho=1$ et x variant de -2 à 2 . On remarque que le signe de $\tilde{s} \cdot \tilde{x}$ est positif ou nul (resp. négatif) si $|x| \leq \rho$ (resp. $|x| > \rho$).

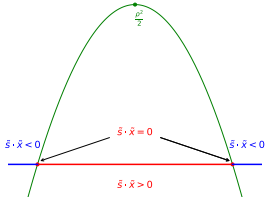


FIGURE 1 : Interprétation de $\tilde{s} \cdot \tilde{x}$ (vert) pour $n=1$.

On appellera couche hypersphérique de largeur m , une concaténation de m neurones hypersphériques. De façon à alléger la notation par la suite, l'indice j désignera toujours le numéro de la sphère alors que l'indice i portera sur la dimension de l'entrée. Ce dernier n'apparaîtra plus en dimension 1.

3 Un théorème d'approximation pour le cas hypersphérique

Dans le cas classique, on peut formuler la sortie d'un réseau de neurones à une couche cachée à m neurones par l'expression suivante :

$$y(\mathbf{x}, \{w_j, b_j, \lambda_j\}_j) = \sum_{j=1}^m \lambda_j \sigma \left(\sum_{i=1}^n w_{j,i} x_i - b_j \right) \quad (2)$$

où σ est une fonction d'activation. De façon analogue à l'équation 2, on peut réécrire la sortie du réseau à couche hypersphérique en

$$y(\mathbf{x}, \{\tilde{s}_j, \lambda_j\}_j) = \sum_{j=1}^m \lambda_j \sigma \left(-\frac{1}{2} \sum_{i=1}^n \left[(\tilde{x}_i - \tilde{s}_{j,i})^2 - \frac{\tilde{s}_j^2}{n} \right] \right) \quad (3)$$

Pour la suite, on considère le problème à une dimension.

La question de densité consiste donc à montrer qu'une combinaison linéaire de sorties de neurones hypersphériques approxime une fonction continue sur un compact de \mathbb{R} . Pour cela, nous nous appuyons sur le théorème suivant de Schwartz :

Théorème 1 [6, p.915] Si Φ et Ψ sont deux fonctions continues à supports compacts, $\Phi \neq 0$, alors Ψ est limite uniforme sur tout compact de combinaisons linéaires des translatées de Φ .

Si l'on note Ω un compact de \mathbb{R} , on peut donc dire qu'il existe pour $j \in \{1, \dots, m\}$, des $(a_j, \lambda_j) \in \mathbb{R}^2$ tels que :

$$\lim_{m \rightarrow +\infty} \sup_{x \in \Omega} \left| \Psi(x) - \sum_{j=1}^m \lambda_j \Phi(x - a_j) \right| = 0 \quad (4)$$

On note $\mathcal{M} := \text{vect}\{\Phi(x - a), a \in \mathbb{R}\}$. Chaque élément de la somme appartenant à \mathcal{M} , la limite de cette somme pour $m \rightarrow +\infty$ est dans l'adhérence $\overline{\mathcal{M}}$ de \mathcal{M} . Cette limite Ψ étant par ailleurs dans $C_c^0(\mathbb{R})$, on en déduit que $C_c^0(\mathbb{R}) \subseteq \overline{\mathcal{M}}$.

Tout élément de \mathcal{M} étant une fonction continue à support compact alors $\mathcal{M} \subseteq C_c^0(\mathbb{R})$. Il suit que $\overline{\mathcal{M}} \subseteq C_c^0(\mathbb{R})$. On a donc montré que \mathcal{M} est dense dans $C_c^0(\mathbb{R})$,

$$\overline{\text{vect}\{\Phi(x - a), a \in \mathbb{R}\}} = C_c^0(\mathbb{R}) \quad (5)$$

Afin d'établir notre théorème d'approximation, nous allons maintenant appliquer cette propriété pour deux cas distincts :

1. Les rayons des sphères sont fixés ($\rho_j = 1$ par exemple).
2. Les centres des sphères sont fixés ($c_j = 0$ par exemple).

Pour les rayons fixés, on introduit la fonction Φ_ρ définie sur \mathbb{R} comme la sortie d'un neurone pour une sphère de centre $c = 0$ et de rayon 1 : $\Phi_\rho(x) = \sigma \left(-\frac{1}{2} [\|x\|^2 - 1] \right)$.

La sortie d'un neurone pour une sphère de centre c et de rayon 1 (cf. eq. 3) s'identifie aisément comme une translatée de Φ_ρ :

$$\Phi_\rho(x - c) = \sigma \left(-\frac{1}{2} [\|x - c\|^2 - 1] \right)$$

Si $\sigma \in C_c^0(\mathbb{R})$ et $\sigma \neq 0$, le théorème 1 s'applique et assure l'existence de $(c_j, \lambda_j) \in \mathbb{R}^2$ tels que :

$$\lim_{m \rightarrow +\infty} \sup_{x \in \Omega} \left| \Psi(x) - \sum_{j=1}^m \lambda_j \sigma \left(-\frac{1}{2} [\|x - c_j\|^2 - 1] \right) \right| = 0$$

ce qui permet de conclure en reprenant un raisonnement analogue à l'obtention de l'équation 5 que :

$$\overline{\mathcal{M}_c} = C_c^0(\mathbb{R}) \quad (6)$$

où $\mathcal{M}_c = \text{vect} \left\{ \sigma \left(-\frac{1}{2} [\|x - c\|^2 - 1] \right) ; c \in \mathbb{R} \right\}$. ■

Pour les centres fixés, on établit un raisonnement similaire en posant les changements de variables suivants :

$$X = -\frac{1}{2}x^2 + \frac{1}{2} \quad \text{et} \quad A = -\frac{1}{2}\rho^2 + \frac{1}{2}$$

On définit également les fonctions

$$\Phi_c(X) = \sigma(X) \quad \text{et} \quad \Psi^*(X) = \Psi(x)$$

Ψ^* est une fonction continue sur le support compact Ω^* , lié à Ω par le changement de variable.

Si l'on effectue la translation de Φ_c par A , à savoir

$$\begin{aligned} \Phi_c(X - A) &= \sigma\left(-\frac{1}{2}x^2 + \frac{1}{2} - \left(-\frac{1}{2}\rho^2 + \frac{1}{2}\right)\right) \\ &= \sigma\left(-\frac{1}{2}(x^2 - \rho^2)\right) \end{aligned}$$

celle-ci correspond à une sortie hypersphérique centrée en 0 et de rayon ρ (cf. eq. 3). Le théorème 1 s'applique encore et donne

$$\lim_{m \rightarrow \infty} \sup_{X \in \Omega^*} \left| \Psi^*(X) - \sum_{j=1}^m \lambda_j \Phi_c(X - A_j) \right| = 0$$

ce qui s'écrit aussi :

$$\lim_{m \rightarrow \infty} \sup_{x \in \Omega} \left| \Psi(x) - \sum_{j=1}^m \lambda_j \sigma\left(-\frac{1}{2}(x^2 - \rho_j^2)\right) \right| = 0$$

Par suite, si on obtient $\overline{\mathcal{M}_\rho} = \mathcal{C}_c^0(\mathbb{R})$ (7)

où $\mathcal{M}_\rho = \text{vect} \left\{ \sigma\left(-\frac{1}{2}[x^2 - \rho^2]\right); \rho \in \mathbb{R}^* \right\}$. ■

D'après les équations 6 et 7, on obtient la suite d'inclusions

$$\overline{\mathcal{M}_\rho} \cup \overline{\mathcal{M}_c} = \mathcal{C}_c^0(\mathbb{R}) \subseteq \overline{\mathcal{M}_{c\rho}}$$

A l'opposé, chaque élément de \mathcal{M}_ρ ou \mathcal{M}_c appartenant à $\mathcal{C}_c^0(\mathbb{R})$, on a donc

$$\mathcal{M}_{c\rho} \subseteq \overline{\mathcal{M}_{c\rho}} \subseteq \mathcal{C}_c^0(\mathbb{R}).$$

Nous aboutissons au théorème d'approximation suivant :

$$\overline{\text{vect} \left\{ \sigma\left(-\frac{1}{2}[\|c - x\|^2 - \rho^2]\right); c \in \mathbb{R}, \rho \in \mathbb{R}^* \right\}} = \mathcal{C}_c^0(\mathbb{R}) \quad (8)$$

4 Expérimentations

Afin de vérifier expérimentalement la conclusion établie dans la partie précédente (équation 8), nous avons testé l'approximation de fonctions continues à support compact à une variable. Pour cela, nous avons utilisé des réseaux de m neurones sur une couche cachée de type hypersphériques ou dense. Pour chaque réseau, nous avons fait varier le nombre de neurones de 2 à 64. Comme fonction de perte, nous avons utilisé l'erreur quadratique moyenne (EQM) et nous avons séparé le jeu de données (de 512 points) selon les proportions suivantes : $\frac{1}{2}$ pour l'entraînement, $\frac{1}{4}$ pour la validation et $\frac{1}{4}$ pour le test. Les données ont été échantillonnées sur l'intervalle $[-0.05, 1.05]$. L'erreur finale est évaluée sur le jeu de test.

Dans nos modèles, nous avons testé puis sélectionné différentes fonctions d'activation que nous avons appliquées

aux neurones de la couche cachée (cf. Fig 2) : linear (sans activation), relu, hard_sigmoid, gauss, tanh, tanhpl, sinus et $gauss_{x\delta}(x) = \left(\frac{\tilde{s} \cdot \tilde{x}}{\delta}\right) e^{-\left(\frac{\tilde{s} \cdot \tilde{x}}{\delta}\right)^2}$ (δ est un paramètre appris).

La figure 2 montre l'évaluation des fonctions d'activation sur le produit interne $\tilde{x} \cdot \tilde{s}$, en considérant une sphère \tilde{s} de centre 0 et de rayon 1 et en faisant varier x sur l'intervalle $[-\pi, \pi]$. En pointillé, sont représentées les fonctions d'activations usuelles. La courbe verte correspondant à l'activation "linear" montre le produit interne $\tilde{s} \cdot \tilde{x}$, c'est-à-dire la sortie du neurone sans activation.

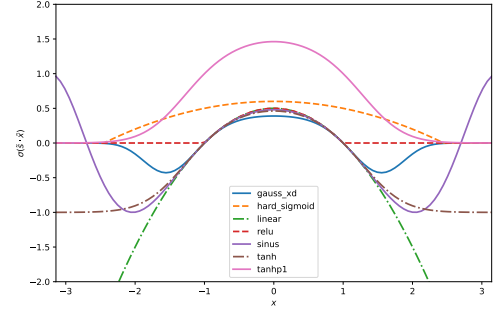


FIGURE 2 : Fonctions d'activation appliquées à $\tilde{s} \cdot \tilde{x}$

Nous avons préalablement établi que la fonction d'activation choisie doit être une fonction à support compact. Dans le but de se rapprocher au plus près des conditions, nous nous intéressons aux fonctions L^2 qui s'annulent au bord. C'est pourquoi nous avons ajouté la fonction $tanhpl(x) := \tanh(x) + 1$. De même, nous avons introduit la fonction $gauss_{x\delta}$ qui possède la propriété de conserver le signe de $\tilde{s} \cdot \tilde{x}$. Enfin, nous avons inclus la fonction sinus comme non linéarité car elle permet de modéliser efficacement des signaux complexes.

Dans un premier temps, nous avons cherché à approximer deux fonctions simples construites pour être à support compact : la fonction Triangle facile à approcher par une couche dense, la fonction Omega qui correspondant à l'activation $relu$ d'une couche à un neurone hypersphérique. Dans un deuxième temps, nous avons étudié deux fonctions plus complexes : la fonction $Sph32$ qui est la combinaison linéaire (+ $relu$) de 32 sphères aléatoires, la fonction $\sin\left(\frac{2\pi}{x}\right)$ qui oscille fortement autour de 0. Chaque fonction à approcher a été mise à 0 en dehors de $[0, 1]$. La figure 3 illustre les fonctions considérées.

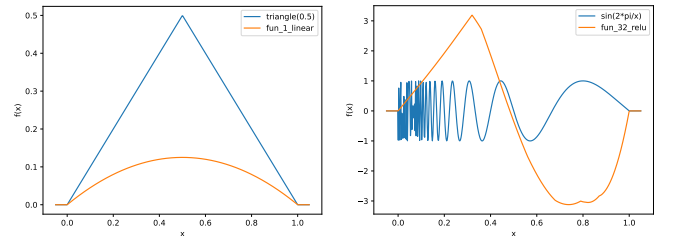


FIGURE 3 : Fonctions à approcher

La figure 4 montre l'approximation des fonctions Triangle et Omega. La légende indique qu'il est nécessaire de monter à 64 neurones pour bien approximer également pour les intervalles $[-0.05, 0]$ et $[1, 1.05]$. On peut noter que le modèle de neurones à couche hypersphérique permet de diminuer les erreurs au bord mais présente quelques oscillations sur les parties purement linéaires de la fonction. Cependant, avec la fonction d'activation $Gauss_{x\delta}$, on obtient une EQM de 8×10^{-6} , qui

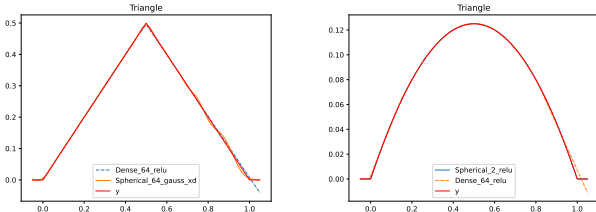


FIGURE 4 : Approximation des fonctions Triangle et Omega

est inférieure à l'EQM maximale obtenue dans le cas des réseaux à couche dense, qui est égal à 2×10^{-5} .

On peut également noter la difficulté à entraîner les modèles à couche hypersphérique. En effet, pour la fonction Omega construite à partir du produit interne avec une hypersphère sur laquelle on compose avec un ReLU, un seul neurone ne suffit pas pour obtenir l'approximation optimale. Il faut au moins deux neurones.

Le graphique 5 représente les EQM, en fonction du nombre de neurones pour les fonctions plus complexes. Les courbes continues représentent les résultats obtenus avec les modèles de neurones à couche hypersphérique, tandis que les courbes en pointillés représentent les résultats obtenus avec les modèles à couche dense.

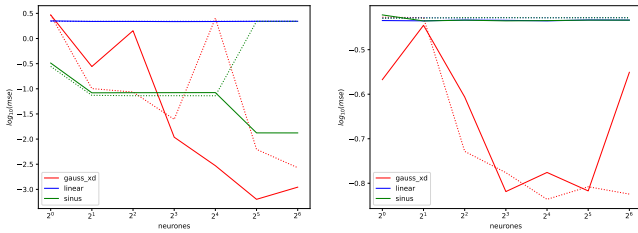


FIGURE 5 : EQM test : $Sph32$ (à g.) et $\sin(2\pi/x)$ (à d.) hypersphérique (resp. dense) en trait plein (resp. pointillé)

Les EQM restent dans les mêmes ordres de grandeur, quel que soit le modèle utilisé. On peut observer que pour l'approximation de $Sph32$, l'utilisation de la fonction d'activation \sinus permet d'obtenir de bons résultats avec peu de neurones. Cependant, que ce soit pour les modèles à couche dense ou hypersphérique, en augmentant suffisamment le nombre de neurones, l'utilisation de la fonction d'activation $Gauss_{x,\delta}$ a permis de réduire les erreurs obtenues pour l'ensemble des fonctions que nous avons approximées.

Nous avons également testé l'approximation de fonctions en fixant le rayon ρ des sphères. Nous n'avons pas noté de changement significatif sur les résultats obtenus, les EQM restant toujours dans les mêmes ordres de grandeur.

5 Discussion et conclusion

Plus généralement, on peut faire le lien entre un réseau à couche hypersphérique (RHS) et un réseau de fonctions de base radiale (RBFN) connu pour être également un approximateur universel [4]. La sortie d'un RBFN s'exprime comme :

$$y(\mathbf{x}, \{c_j, \gamma_j, \lambda_j\}_j) = \sum_{j=1}^m \lambda_j \phi(\gamma_j \|\mathbf{x} - c_j\|) \quad (9)$$

La fonction ϕ est dite à base radiale car elle ne dépend que de la distance euclidienne entre \mathbf{x} et $c_j \in \mathbb{R}^n$. Pour rapprocher les sorties des deux types réseaux, les sphères doivent être dégénérées ($\rho_j = 0$) en points. Dans ce cas, la sortie d'un RHS pour l'activation $\sigma(x) = x$ (resp. $\sigma(x) = e^x$) est proportionnelle à la sortie d'un RBFN en utilisant le noyau quadratique $\phi(x) = x^2$ (resp. le noyau $\phi(x) = e^{-\frac{1}{2}x^2}$).

Concernant les conditions de l'approximation universelle, ϕ doit être $L^1(\mathbb{R}^n)$ dans un RBFN et vérifier $\int_{\mathbb{R}^n} \phi(x) d(x) \neq 0$ alors que la fonction d'activation σ doit vérifier $\sigma \in C_c^0(\mathbb{R})$ dans un RHS. La mise en oeuvre de ces deux types de réseaux s'avèrent toutefois très différentes par rapport à l'initialisation (clustering vs Glorot) et l'optimisation des paramètres. Les RHS sont généralement difficiles à entraîner lorsque la dimension augmente car dépendent fortement de l'initialisation. Nous travaillons actuellement sur ce point.

Les résultats expérimentaux permettent de confirmer le résultat théorique suivant : un réseau de neurones à une couche hypersphérique est capable d'approcher en norme uniforme sur tout compact toute fonction continue de \mathbb{R} à support compact. Nous avons proposé une nouvelle fonction d'activation $Gauss_{x,\delta}$ qui semble prometteuse dans le modèle hypersphérique et dense. Une étude comparative avec la fonction d'activation \sinus est une perspective. Expérimentalement, nous avons vérifié que l'approximation est malgré tout possible par \tanh et \sinus même si elles ne sont pas à support compact (mais $L^2(\Omega)$). Cela soulève des questions sur une approche différente du théorème d'approximation universelle. Nous envisageons d'effectuer une approche similaire à celle présentée par A. Pinkus [5], afin de réduire les hypothèses nécessaires sur les fonctions à approximer et les fonctions d'activation, et de généraliser le théorème d'approximation aux fonctions à plusieurs variables. Au delà de la possibilité d'approximer une fonction, nous envisageons d'utiliser des sphères et les quantités signées \tilde{s}_j, \tilde{x} dans le contexte de la détection d'anomalie.

Références

- [1] Vladimir BANARER, Christian PERWASS et Gerald SOMMER : The hypersphere neuron. *In ESANN*, pages 469–474, 2003.
- [2] Julien de SAINT ANGEL et Christophe SAINT-JEAN : Couches Dense et Conv2d “ sphériques ” par l'algèbre géométrique conforme. *In ORASIS 2021*.
- [3] Hongbo LI, David HESTENES et Alyn ROCKWOOD : *Spherical Conformal Geometry with Geometric Algebra*, pages 61–75. Springer Berlin Heidelberg, 2001.
- [4] Jooyoung PARK et Irwin W. SANDBERG : Universal approximation using radial-basis-function networks. *Neural Computation*, 3:246–257, 1991.
- [5] Allan PINKUS : Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1999.
- [6] Laurent SCHWARTZ : Théorie générale des fonctions moyenne-périodiques. *Annals of Mathematics*, 48(4):857–929, 1947.